



MONTCAS, PHASE 2 Criterion-Referenced Test

2004 TECHNICAL MANUAL

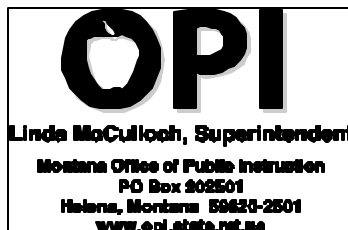


TABLE OF CONTENTS

SECTION I: ASSESSMENT DEVELOPMENT.....	5
CHAPTER 1 – BACKGROUND AND OVERVIEW.....	5
PURPOSE OF THIS MANUAL.....	5
PURPOSE OF THE CRT.....	5
OPTIONS FOR PARTICIPATION.....	6
ORGANIZATION OF THIS MANUAL.....	6
CHAPTER 2—OVERVIEW OF TEST DESIGN.....	7
CRT.....	7
ITEM TYPES.....	7
COMMON-MATRIX DESIGN.....	8
TEST SCHEDULING.....	8
CHAPTER 3 – TEST DEVELOPMENT PROCESS.....	10
PROGRESS TOWARDS STANDARDS ITEM DEVELOPMENT.....	10
PROGRESS TOWARDS STANDARDS ITEM DEVELOPMENT PROCESS OVERVIEW.....	11
MONTANA-AUGMENTED ITEM DEVELOPMENT.....	12
MONTANA-AUGMENTED ITEM DEVELOPMENT PROCESS OVERVIEW.....	13
INTERNAL ITEM REVIEW.....	14
EXTERNAL ITEM AND BIAS REVIEWS.....	14
ITEM EDITING.....	15
OPERATIONAL TEST ASSEMBLY.....	15
EDITING DRAFTS OF OPERATIONAL TESTS.....	16
BRAILLE AND LARGE-PRINT TRANSLATION.....	17
CHAPTER 4 - DESIGN OF THE READING ASSESSMENT.....	18
READING BLUEPRINT.....	18
ITEM TYPES.....	19
TEST DESIGN.....	19
CHAPTER 5—DESIGN OF THE MATHEMATICS ASSESSMENT.....	20
MATHEMATICS BLUEPRINT.....	20
CONTENT SPECS.....	20
ITEM TYPES.....	21
TEST DESIGN.....	22
THE USE OF CALCULATORS IN THE CRT.....	22
SECTION II: TEST ADMINISTRATION.....	23
CHAPTER 6—TEST ADMINISTRATION.....	23
RESPONSIBILITY FOR ADMINISTRATION.....	23
PROCEDURES.....	23
ADMINISTRATOR TRAINING.....	23
PARTICIPATION REQUIREMENTS.....	24
SECTION III: DEVELOPMENT AND REPORTING OF SCORES.....	25
CHAPTER 7—SCORING.....	25
MACHINE-SCORED ITEMS.....	25
SCANNING QUALITY CONTROL.....	26
ELECTRONIC DATA FILES.....	26
ITEMS SCORED BY READERS.....	27
I-SCORE.....	27
PRELIMINARY ACTIVITIES.....	28
PLANNING AND DESIGNING DOCUMENTS.....	28
BENCHMARKING.....	29
SELECTING AND TRAINING SCORING STAFF.....	29
QUALITY ASSURANCE COORDINATORS (QACS) AND VERIFIERS.....	29
TRAINING QUALITY ASSURANCE COORDINATORS AND VERIFIERS.....	29
SELECTING READERS.....	29
TRAINING READERS.....	30
SCORING ACTIVITIES.....	30
MONITORING READERS.....	31

GENERAL SCORING GUIDES	31
SHORT-ANSWER ITEMS.....	31
CONSTRUCTED- RESPONSE ITEMS.....	31
CHAPTER 8—EQUATING.....	32
GENERAL RULES.....	32
MATH EQUATING DESIGN.....	32
READING EQUATING DESIGN.....	33
CHAPTER 9—ITEM ANALYSES.....	35
DIFFICULTY INDICES (P).....	35
ITEM-TEST CORRELATIONS (ITEM DISCRIMINATION).....	36
SUMMARY OF ITEM ANALYSIS RESULTS.....	37
DIFFERENTIAL ITEM FUNCTIONING (DIF).....	38
ITEM RESPONSE THEORY ANALYSES.....	40
CHAPTER 10—RELIABILITY.....	42
RELIABILITY AND STANDARD ERRORS OF MEASUREMENT.....	44
CHAPTER 11—SCALED SCORES.....	46
CHAPTER 12—STANDARD SETTING.....	48
CHAPTER 13? REPORTING.....	55
SECTION IV—REFERENCES.....	57
SECTION V—APPENDICES.....	59
APPENDIX A: PTS ITEM AND BIAS REVIEW COMMITTEES & GUIDELINES FOR PTS READING PASSAGES AND ITEM BIAS AND SENSITIVITY REVIEW.....	61
APPENDIX B: ITEM PARAMETER FILES.....	75
APPENDIX C: STANDARD SETTING: EVALUATION SUMMARIES.....	97
APPENDIX D: TECHNICAL ADVISORY COMMITTEE.....	117
APPENDIX E: CRT PERFORMANCE LEVEL DESCRIPTORS, SCALED SCORES, AND RAW SCORES.....	119
APPENDIX F: REPORT SHELLS.....	123
APPENDIX G: REPORTING DECISION RULES.....	129

SECTION I: ASSESSMENT DEVELOPMENT

CHAPTER 1—BACKGROUND AND OVERVIEW

PURPOSE OF THIS MANUAL

The purpose of this manual is to document the technical aspect of the 2004 MontCAS, Phase 2 Criterion-Referenced Test (CRT). In the spring of 2004, students in grades 4, 8, and 10 participated in the administration of the CRT; during this administration reading and mathematics were assessed. This report provides information about the technical quality of those assessments, including a description of the processes used to develop, administer, and score the tests and to analyze the test results.

While some parts of this technical report may be used by educated laypersons, the intended audience is experts in psychometrics and educational research. The report assumes a working knowledge of measurement concepts such as reliability and validity, and statistical concepts such as correlation and central tendency. In some chapters, the reader is presumed to also have basic familiarity with advanced topics in measurement and statistics.

PURPOSE OF THE CRT

On April 5, 2002, the Montana Office of Public Instruction (OPI) entered into a compliance agreement with the U.S. Department of Education that required Montana to implement a number of actions by April 5, 2005, to bring the state into compliance with the provisions of the following federal laws: Title 1 of the Elementary and Secondary Education Act (ESEA) of 1994, P.L. 103-382 and the No Child Left Behind Act (NCLB) of 2001. Montana received federal appropriations to develop an appropriate assessment. The criterion-referenced test (CRT) was developed in accordance with the compliance agreement and federal laws.

The CRTs are based-on, and aligned to, Montana's Content Standards in Reading and Mathematics. Montana educators worked with OPI and its contractor, Measured Progress, in the development and review (content and bias) of these tests to assess how well students have learned the Montana content standards for their grade. The United States Department of Education (USDOE) approved the CRT

assessments in reading and mathematics for grades 3-8 and 10 by school year 2005-2006 and in science at one grade in each of three grade spans (e.g., four, eight, and ten) by school year 2007-2008.

OPTIONS FOR PARTICIPATION

To ensure that all students can participate, the CRT provides accommodations, standard and non-standard, and an alternate assessment.

CRT without accommodations	CRT with standard accommodations	CRT with non-standard accommodations	CRT-Alternate
<ul style="list-style-type: none"> • For students who do not require accommodations of any kind. • Group setting • Untimed – with guidelines 	<ul style="list-style-type: none"> • Available for any student (students with disabilities <u>as well as</u> student without disabilities) when an accommodation is necessary to allow the student to demonstrate his/her skills and competencies. • Must be coded in the Student Response Booklet. • May be given in either, or both, reading and math portions. • Does not change intent/content of the test. 	<ul style="list-style-type: none"> • For a student when specified in his/her IEP/504/LEP plan. • Must be coded in the Student Response Booklet. • May be given in either, or both, reading and math portions. • Changes the intent/content of the test. • When a non-standard accommodation is used, the student's score will be reported in the NOVICE performance category for that content area. 	<ul style="list-style-type: none"> • For a student when specified in his/her IEP plan. • For students who have a significant cognitive disability. • Must be coded in the Student Response Booklet. • May be given in either, or both, reading and math portions. • Based on alternate achievement standards.

ORGANIZATION OF THIS MANUAL

The organization of this manual is based on the conceptual flow of an assessment's life span: it begins with the initial test specification and addresses all the intermediate steps that lead to final score reporting. Section I covers the development of the CRT tests. It consists of five chapters, covering general design issues, the test development process, and the specific designs of the reading and mathematics assessments. Section II consists of a single chapter, Chapter 6, describing the administration of the tests. Section III contains seven chapters, covering scoring, equating, item analysis, reliability and validity, scaling, standard setting, and reporting. We have also included two additional sections: Section IV contains references and Section V contains the appendices.

CHAPTER 2—OVERVIEW OF TEST DESIGN

CRT

CRT questions are directly linked to **Montana’s Content Standards**. The content standards are the basis for the reporting categories developed for each subject area and are used to help guide the development of test questions. No other content or process is subject to statewide assessment. An item may address part, all, or several of the benchmarks within a standard.

ITEM TYPES

Montana’s educators and students were familiar with most of the question types that were used in the assessment program, although one new type—the constructed-response question—was used as well. The types of questions used and the functions of each are described below.

Multiple-choice questions were used, in part, to provide breadth of coverage of a subject area. Because they require no more than a minute for most students to answer, these questions make efficient use of limited testing time and allow coverage of a wide range of knowledge and skills.

Short-answer questions were used to assess students’ skills and their abilities to work with brief, well-structured problems that had one or a very limited number of solutions (e.g., mathematical computations). Short-answer questions require approximately two minutes for most students to answer. The advantage of this type of question is that it requires students to demonstrate knowledge and skills by generating, rather than merely selecting, an answer.

Constructed-response questions typically require students to use higher-order thinking skills—evaluation, analysis, summarization, and so on—in constructing a satisfactory response. Constructed-response questions should take most students approximately five to ten minutes to complete. It should be noted that the use of released CRT questions to prepare students to answer this kind of question is appropriate and encouraged.

COMMON-MATRIX DESIGN

The CRT measures what students know and are able to do by using a greater variety of question types. The tests are structured using both **common** and **matrix-sampled** questions. Common questions are those taken by all students at a given grade level. In addition, a larger pool of matrix-sampled questions is divided among the sixteen forms of the test at each grade level. Each student takes only one form of the test and so answers a fraction of the matrix-sampled questions in the entire pool. This design provides reliable and valid results at the student level.

The CRT reports delivered to schools in September 2004, reported out common scores in the results for ease of understanding them. In addition, common items were released with a data management tool called *iAnalyze*.

TEST SCHEDULING

The CRTs were given during the spring: **reading** and **mathematics** were administered to grades 4, 8 and 10 during a three-week period (March 29 – April 16, 2004). Schools were able to schedule testing sessions at any time during this period, provided they followed the sequence in the scheduling guidelines detailed in test administration manuals. Schools were asked to schedule makeup testing of students who were absent from initial test sessions during this testing window.

The CRT is an un-timed assessment; however, guidelines were provided in the *Test Coordinator's* and *Test Administrator's Manuals* based on estimates of the time it would take an average student to respond to each type of question that made up the test:

- multiple-choice questions – 1 minute per question
- short-answer questions – 2 minutes per question
- constructed-response questions – 10 minutes per question

While the guidelines for scheduling were based on the assumption that most students would complete the test within the time estimated, each test administrator was asked to allow additional time for students who needed it. If additional classroom space was not available for students who required additional time to complete the tests, schools were encouraged to consider using another space, such as the guidance office, for this purpose. If additional areas were not available, it was recommended that each classroom being used for test administration be scheduled for the maximum amount of time.

CHAPTER 3—TEST DEVELOPMENT PROCESS

PROGRESS TOWARD STANDARDS ITEM DEVELOPMENT

The items developed for the Progress Toward Standards common and matrix forms were consistent with the Progress Toward Standards and national content standards. Measured Progress development specialists then aligned the items to the appropriate Montana Content Standards. As an additional quality control check, lead developers in each content area and Montana educators verified that each item was appropriately aligned. In June 2002, Northwest Regional Educational Laboratory (NWREL) performed an independent alignment study to verify item alignment to Montana Content Standards.

The development process Measured Progress followed combined the expertise of the item development team and collaboration with a panel of educators nationwide to help ensure that these items met the needs of the core Progress Toward Standards program and the CRT program. All items used in the Progress Toward Standards common and matrix portions of the CRT program underwent review by a national panel of content and bias reviewers. This panel included numerous Montana educators (see Appendix “A”: Guidelines for Progress Toward Standards Reading Passages & Item Bias and Sensitivity Review).

TABLE 1: TOTAL NUMBER OF PROGRESS TOWARD STANDARDS ITEMS DEVELOPED PER YEAR

GRADE	READING	MATH
4	168	80
8	168	80
10	168	80

PROGRESS TOWARD STANDARDS ITEM DEVELOPMENT PROCESS OVERVIEW

An overview of the test development process for the Progress Toward Standards common and matrix items, including conducting the field tests, follows.

TABLE 2: DEVELOPMENT PROCESS OVERVIEW

DEVELOPMENT STEP	OBJECTIVE OF THE STEP
Develop reading and mathematics content standards	<ul style="list-style-type: none"> Measured Progress Curriculum and Assessment Specialists in the reading area reviewed the guidelines of national curriculum organizations such as NCTE and IRA to determine what students should know and be able to do in reading at each grade level. Measured Progress Curriculum and Assessment Specialists in the reading area reviewed the academic content standards of states nationwide to help ensure that the content standards for the Progress Toward Standards tests would be consistent with the content standards of most states. Measured Progress Curriculum and Assessment Specialists in the reading and mathematics developed draft content standards. The draft content standards were reviewed by a panel of external reading experts. Measured Progress Curriculum and Assessment Specialists revised the content standards according to feedback from the external review and finalized the document.
Develop common and matrix test blueprints	<ul style="list-style-type: none"> Measured Progress senior technical advisors and psychometricians developed test specifications for the common forms to help ensure adequate reliability, validity and fairness of the tests for the purposes of large-scale, high-stakes assessment. Measured Progress senior technical advisors and psychometricians developed test specifications for the matrix forms to help ensure that sufficient items are field tested so that newly-developed, high quality and comparable common test forms are available each year.
National educators reviewed passages and reading and math items (December 2002; March 2003; December 2003)	<ul style="list-style-type: none"> Measured Progress Curriculum and Assessment Specialists in the reading area located potential reading passages. National educators reviewed reading and mathematics items.
Alignment Study (June 2002)	<ul style="list-style-type: none"> Northwest Regional Educational Laboratory performed an independent alignment study to verify that all items (PTS and MT Augmented) correlate to Montana's Content Standards.
Item and Bias Review Meetings (Sept. 2002)	Curriculum and Assessment Specialists facilitated the review of all items and selected appropriate items for field testing.

National Field Tests of Items (Fall 2002 and Spring 2003)	Measured Progress administered a national field test to a sample of students in seven states prior to use of the items in operational assessment to assure quality of items.
Final Item Selection (August 2003)	Measured Progress provided the reports necessary for national educators to review the results of field-testing, revise as necessary, and select items for the assessment.

MONTANA-AUGMENTED ITEM DEVELOPMENT

The items developed for the augmented CRT were consistent with Montana's content standards. Using a consensus model, our development specialists worked with OPI and Montana educators to align the items developed to augment the CRT to appropriate Montana content standards. As an additional quality control check, lead developers in each content area checked for their agreement that each item was appropriately aligned. Where there were any questions or apparent discrepancies, our lead Curriculum and Assessment Specialists resolved them with OPI.

The development process Measured Progress followed, combining the expertise of the item development team and collaboration with Montana educators, helping to ensure that these items met the needs of the CRT program. The item specifications were built on the Montana content standards, thus assuring complete alignment between the content standards and the augmented portion of the CRT. In addition to internal review, all test materials and items used in the CRT program underwent review by Montana educators and bias review committees prior to print.

TOTAL NUMBERS OF MONTANA-AUGMENTED ITEMS DEVELOPED BY GRADE AND CONTENT

GRADE	READING	MATH
3	60	60
4	100	100
5	60	60
6	60	60
7	60	60
8	100	100
10	150	150

MONTANA-AUGMENTED ITEM DEVELOPMENT PROCESS OVERVIEW

An overview of the test development process for the CRT program, including conducting the pilot tests (April 2003), follows.

DEVELOPMENT PROCESS OVERVIEW

DEVELOPMENT STEP	OBJECTIVE OF THE STEP
Montana educators reviewed passages for the reading tests (Aug. 2002)	<ul style="list-style-type: none"> Measured Progress Curriculum and Assessment Specialists in the reading area located potential reading passages. MT educators approved the passages prior to item writing in consultation with a Montana Bias Review Committee. Measured Progress Permissions staff secured permissions to use the passages prior to item writing meetings.
Item drafting/editing meetings (Sept. 2002)	Measured Progress Curriculum and Assessment Specialists - <ul style="list-style-type: none"> provided item development training to Montana participants. facilitated the development of item ideas by the participants.
Editorial review of items (Oct. 2002)	All items were reviewed by members of Measured Progress's Publications staff to assure - <ul style="list-style-type: none"> clarity and unambiguousness of items. correct grammar, punctuation, usage, and spelling. technical quality with respect to stems, options, and scoring guides. compliance with OPI sensitivity standards and style guidelines.
Item review meetings (Nov. 2002)	Curriculum and Assessment Specialists facilitated the review of all items with Montana educators and selected appropriate items for field testing in 2003.
Bias Review Committee meetings (Nov. 2002)	Measured Progress staff facilitated the review of all test items for sensitivity and bias considerations based on OPI guidelines. Members of this committee were selected by OPI. Measured Progress provided OPI with guidelines for committee membership.
Field Test of MT-Augmented Items (April 2003)	Measured Progress provided field test forms which were administered to a sample of students in Montana prior to use of the items in operational assessment to assure quality of items.
Final Item Selection (August 2003)	Measured Progress provided the reports necessary for Montana educators to review the results of field-testing, revise as necessary, and select items for the augmented portion of the assessment.

INTERNAL ITEM REVIEW

- The lead or peer test developer within the content specialty reviewed the typed item, the constructed-response scoring guide, and any reading selections and graphics.
- The content reviewer considered item “integrity”; item content and structure; appropriateness to designated content area; item format; clarity; possible ambiguity; keyability; single “keyness”; appropriateness and quality of reading selections and graphics; and appropriateness of scoring guide descriptions and distinctions (as correlated to the item and within the guide itself).
- The content reviewer also considered scorability and evaluated whether the scoring guide adequately addressed performance on the item.
- Fundamental questions the content reviewer considered, but was not limited to, included the following:
 - What is the item asking?
 - Is the key the only possible key?
 - Is the constructed-response item scorable as written (are the correct words used to elicit the response defined by the guide)?
 - Is the wording of the scoring guide appropriate and parallel to the item wording?
 - Is the item complete (e.g., with scoring guide, content codes, key, grade level, and contract identified)?
 - Is the item appropriate for the designated grade level?

EXTERNAL ITEM AND BIAS REVIEWS

- Common item sets were delivered to OPI for Montana educator content and bias reviews.
- The PTS National Bias and Content Review Committee reviewed the common and matrix passages and items used for the 2004 administration in Montana during two two-day meetings, held in March 2003 and December 2003 in Chicago, IL, and during a mail review of passages in July 2003 (see Appendix “A”).

ITEM EDITING

Editors reviewed and edited the items to ensure uniform style (based on *The Chicago Manual of Style*) and adherence to sound testing principles. These principles included the stipulation that items

- were correct with regard to grammar, punctuation, usage, and spelling;
- were written in a clear, concise style;
- contained unambiguous explanations for students as to what was required to attain a maximum score;
- were written at a reading level that would allow the student to demonstrate his or her knowledge of the tested subject matter regardless of reading ability;
- exhibited high technical quality regarding psychometric characteristics;
- had appropriate answer options or score-point descriptors; and
- were free of potentially insensitive content.

OPERATIONAL TEST ASSEMBLY

Test assembly is the sorting and laying out of item sets into test forms. Criteria considered during this process included the following:

- Content coverage/match to test design. The curriculum specialist completed an initial sorting of items into sets based on a balance of content categories across sessions and forms, as well as a match to the test design (e.g., number of multiple-choice, short-answer, and constructed-response items).
- Item difficulty and complexity. Item statistics drawn from the data analysis of previously tested items were used to ensure that there were similar levels of difficulty and complexity across forms.
- Visual balance. Item sets were reviewed to ensure that each reflected a similar length and “density” of selected items (e.g., length/complexity of reading selections or number of graphics).
- Option balance. Each item set was checked to verify that it contained a roughly equivalent number of key options (As, Bs, Cs, and Ds).
- Name balance. Item sets were reviewed to ensure that a diversity of names was used.
- Bias. Each item set was reviewed to ensure fairness and balance based on gender, ethnicity, religion, socioeconomic status, and other factors.

- Page fit. Item placement was modified to ensure the best fit and arrangement of items on any given page.
- Facing-page issues. For multiple items associated with a single stimulus (a graphic or a reading selection), consideration was given to whether those items needed to begin on a left- or right-hand page, as well as to the nature and the amount of material that needed to be placed on facing pages. These considerations served to minimize the amount of page flipping required of the students.
- Relationships between forms. Sets of common items were placed identically in each version of the forms. Although matrix-sampled item sets differed from form to form, they took up the same number of pages in each form so that sessions and content areas began on the same page in every form. Therefore, the number of pages needed for the longest form often determined the layout of each form.
- Visual appeal. The visual accessibility of each page of the form was always taken into consideration, including such aspects as the amount of white space, the density of the text, and the number of graphics.

EDITING DRAFTS OF OPERATIONAL TESTS

Any changes made by the test construction specialist had to be reviewed and approved by the test developer. Once a form had been laid out in what was considered its final form, it was reread to identify any final considerations, including the following:

- Editorial changes. All text was scrutinized for editorial accuracy, including consistency of instructional language, grammar, spelling, punctuation, and layout. Measured Progress's publishing standards are based on *The Chicago Manual of Style*.
- Keying items. Items were reviewed for any information that might "key" or provide information that would help students answer another item. Decisions about moving keying items were based on the severity of the key-in and the placement of the items in relation to each other within the form.
- Key patterns. The final sequence of keys was reviewed to ensure that the order appeared random (i.e., no recognizable pattern and no more than three of the same key in a row).

BRaille AND LARGE-PRINT TRANSLATION

Form one for grades 4, 8, and 10 tests was translated into Braille by National Braille Press, a subcontractor, who specializes in test materials for blind and visually handicapped students. In addition, form one for each grade was adapted into a large-print version.

CHAPTER 4—DESIGN OF THE READING ASSESSMENT

READING BLUEPRINT

As indicated earlier, the framework for reading was based on Montana’s standards, which identifies five **content standards** that apply specifically to reading and reading comprehension. Those content standards are

- **Reading Standard 1:** Students construct meaning as they comprehend, interpret, and respond to what they read.
- **Reading Standard 2:** Students apply a range of skills and strategies to read.
- **Reading Standard 3:** Students set goals, monitor, and evaluate their reading progress. (Cannot measure this benchmark with traditional paper/pencil test.)
- **Reading Standard 4:** Students select, read, and respond to print and nonprint material for a variety of purposes.
- **Reading Standard 5:** Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences.

Passages included both long and short texts selected from reading sources that students at each grade level would be likely to encounter in their classroom and in their independent reading. None of the passages were written specifically for the assessment, but instead were collected from published works.

- **Literary passages** are represented by a variety of genres—modern narratives; diary entries; drama; poetry; biographies; essays; excerpts from novels; short stories; and traditional narratives, such as fables, myths, and folktales.
- **Content passages** are primarily informational and often deal with the areas of science and social studies. They are drawn from such sources as newspapers, magazines, and books.
- **Practical passages** are functional materials that instruct or advise the reader—for example, directions, reference tools, or manuals.

The main difference in the passages used for grades 4, 8, and 10 was their degree of difficulty. All passages were selected to be appropriate for the intended audience; however, the ideas expressed became increasingly more complex at grade levels 8 and 10.

The questions related to these passages required students to demonstrate their skills in both literal comprehension (where the answer is stated explicitly in the text) and inferential comprehension (where the answer is implied by the text and/or the text must be connected to relevant prior knowledge to determine an answer). In addition, some questions focused on the reading skills reflected in content standards. Questions of this type required students to use the skills and strategies of reading to answer questions—for example, how to identify the author’s principal purpose, such as to persuade, entertain, or inform—and to demonstrate their understanding of how words and images communicate to readers.

ITEM TYPES

The CRT assessment in reading included multiple-choice and constructed-response questions. Constructed-response questions required students to write an answer consisting of several phrases or short sentences. Each type of question was worth a specific number of points in the student’s total language arts score, as shown below.

Type of Question	Possible Score Points
Multiple-Choice	0–1
Constructed-Response	0–4

TEST DESIGN

The table below summarizes the number and types of questions that were used in the CRT reading assessment for 2004.

		Common Reading Items		TOTAL	
Grade	Session 1	Session 2	Session 3	MC	CRs
4	21 MC, 1 CR	10 MC	21 MC, 1 CR	52	2
8	21 MC, 1 CR	10 MC	21 MC, 1 CR	52	2
10	21 MC, 1 CR	15 MC	21 MC, 1 CR	57	2

Key

- MC = multiple-choice questions
- CR = constructed-response questions

CHAPTER 5—DESIGN OF THE MATHEMATICS ASSESSMENT

MATHEMATICS BLUEPRINT

The mathematics framework was based on Montana’s Mathematics Content Standards, which identifies seven **content standards**, as shown below:

- **Mathematics Standard 1:** Problem Solving
- **Mathematics Standard 2:** Numbers and operations
- **Mathematics Standard 3:** Algebra
- **Mathematics Standard 4:** Geometry
- **Mathematics Standard 5:** Measurement
- **Mathematics Standard 6:** Data Analysis, Statistics, and Probability
- **Mathematics Standard 7:** Patterns, Relations, and Functions

CONTENT SPECS

For students to function effectively as mathematical problem solvers, they must be taught how to apply and communicate basic concepts and procedures, as well as how to do the procedures themselves.

Content questions measure what students have been taught directly. Included in these are the basic concepts and procedural skills from all the content standards. For example, in the numbers and number sense standard and the computation standard, conceptual and procedural knowledge includes understanding of place value in our number system; the computational algorithms as applied to whole numbers, fractions, and decimals; and the concepts of ratio, proportion, and percent. In the data analysis and statistics standard, conceptual and procedural knowledge includes the reading of charts and graphs as well as the concepts of averages (means, medians, and modes) and the methods for computing them. Contextual settings used in questions measuring this category were very simple and were directly related to those used in the teaching of the concepts and the procedures.

Application questions measure what the students can do with what they have been taught. Included are questions requiring students to combine the basic concepts and procedures to solve real-life and mathematical problems, to evaluate their own ideas and the ideas of others using mathematical reasoning, and to communicate their ideas using the wealth of symbolic, pictorial, graphic, and verbal representations available in mathematics.

It is important to understand that application questions also measure mastery of the basic concepts and procedures. For example, in mathematics, questions were either short-answer or constructed-response questions (see “Item Types” in the table below), which were worth up to four score points. In most cases, portions of these questions required the student to perform some problem solving, reasoning, and/or communicating. At the same time, however, the questions required the students to demonstrate their understanding of mathematics content. If a student did not show mastery of all aspects of a constructed-response question, or if he/she made careless errors, the student did not earn the highest score for that question. Thus, it can be said that **all** mathematics questions in the CRT measured content; some questions went beyond that realm (short-answer and constructed-response), however, and were classified as application.

ITEM TYPES

The CRT mathematics assessment included multiple-choice, short-answer, and constructed-response questions. Short-answer questions required students to perform a computation or solve a simple problem. Constructed-response questions were more complex, requiring 8-10 minutes of response time. Each type of question was worth a specific number of points in the student’s total mathematics score, as shown below.

Type of Question	Possible Score Points
Multiple-Choice	0–1
Short-Answer	0–1
Constructed-Response	0–4

TEST DESIGN

The table below summarizes the numbers and types of questions that were used in the CRT, and shows the construction of the common portions of the assessment.

	Common Math Items				TOTAL	
Grade	Session 1 Cal	Session 2A Cal	Session 2B No Cal	Session 3 No Cal	MC	SA & CRs
4	24 MC, 1 CR	5 MC	5 MC	21 MC, 3 SA, 1 CR	55	3 SA, 2 CRs
8	24 MC, 1 CR	5 MC	5 MC	21 MC, 3 SA, 1 CR	55	3 SA, 2 CRs
10	24 MC, 1 CR	8 MC	7 MC	21 MC, 3 SA, 1 CR	60	3 SA, 2 CRs

Key

- Cal = calculator use allowed
- No Cal = no calculator use allowed
- MC = multiple-choice questions
- SA = short-answer questions
- CR = constructed-response questions

THE USE OF CALCULATORS IN THE CRT

The Montana educators who helped develop the CRT acknowledged the importance of mastering arithmetic algorithms. At the same time, they understood that the use of calculators is a necessary and important skill in society today. Calculators can save time and prevent error in the measurement of some higher-order thinking skills and allow students to do more sophisticated and intricate problems. For these reasons, calculators were permitted on some parts of the CRT mathematics assessment and prohibited on others. (Students were allowed to use any calculator with which they were familiar.)

SECTION II: TEST ADMINISTRATION

CHAPTER 6—TEST ADMINISTRATION

RESPONSIBILITY FOR ADMINISTRATION

As indicated in the *Test Coordinator's Manual*, principals and/or their designated School Test Coordinators were responsible for the proper administration of the CRT. Manuals and Principal Certification Forms were used to ensure the uniformity of administration procedures from school to school.

PROCEDURES

School Test Coordinators were instructed to read the *Test Coordinator's Manual* prior to testing, and to be familiar with the instructions given in the *Test Administrator's Manual*. The *Test Coordinator's Manual* provided each school with checklists to help it prepare for testing. The checklists outlined tasks for the schools to perform before, during, and after test administration. Along with these checklists, the *Test Coordinator's Manual* outlined the nature of the testing material being sent to each school, how to inventory the material, how to track it during administration, and how to return the material once testing was complete. It also contained information about including or excluding students. The *Test Administrator's Manual* included checklists for the administrators to prepare themselves, their classrooms, and their students for the administration of the test. The *Test Administrator's Manual* contained sections that detailed the procedure to be followed for each test session, and it contained instructions on preparing the material prior to giving it to the School Test Coordinator for its return to Measured Progress.

ADMINISTRATOR TRAINING

In addition to distributing the *Test Coordinator's* and *Test Administrator's Manuals*, OPI along with Measured Progress conducted pre-administration workshops (one MetNet and one videostream) to train and inform school personnel about the new CRT. In addition, videotapes of the workshops were mailed to System Test Coordinators for distribution in their schools.

PARTICIPATION REQUIREMENTS

The following categories of students were allowed to be considered for exclusions :

- Foreign exchange students
- Students not enrolled (home-schooled)
- Students not in school a full academic year
- Students not in district a full academic year

Information about the exclusion was coded in by staff after testing was completed. The *Test Coordinator's* and *Test Administrator's Manuals* provided directions on coding. Please refer to Appendix G: Decision Rules regarding reporting exclusions.

SECTION III:

DEVELOPMENT AND REPORTING OF SCORES

CHAPTER 7—SCORING

MACHINE-SCORED ITEMS

Once the 2004 test booklets had been logged in, identified with appropriate scannable, preprinted school information sheets, examined for extraneous materials, and batched, they were moved into the scanning area. For all student response booklets (and other forms that required imaging/scanning) this area was the last step in the processing loop in which the documents themselves were handled.

At that point, 100 percent of the student response documents and other scannable information necessary to produce the required reports had been captured and converted into an electronic format, including all student identification and demographics, and digital image clips of short-answer and constructed-response student responses. The digital image clip information allowed Measured Progress to replicate student responses on the readers' monitors just as they had appeared on the originals. From that point on, the entire process—data processing, scoring, benchmarking data analysis, and reporting—was accomplished without further reference to the originals.

The first step in that conversion was the removal of the booklet bindings so that the individual pages could pass through the scanners one at a time. Once cut, the sheets were put back in their proper boxes and placed in storage until needed for the scanning/imaging process.

Customized scanning programs for all scannables were prepared to selectively read the student response booklets and to format the scanned information electronically according to predetermined requirements. Any information (including multiple-choice response data) that had been designated time-critical or process-critical was handled first.

In addition to numerous real-time quality control checks, duplex read, a transport printer that prints a unique identifying number on each sheet of each booklet, and on-line editing capability, the 5000i scanners offer features that make them compatible with Internet technology.

SCANNING QUALITY CONTROL

NCS scanners are equipped with many built-in safeguards that prevent data errors. The scanning hardware is continually monitored for conditions that will cause the machine to shut down if standards are not met. It will display an error message and prevent further scanning until the condition is corrected. The areas monitored include document page and integrity checks, user-designed on-line edits, and many internal checks of electronic functions.

Before every scanning shift begins, Measured Progress operators perform a daily diagnostic routine. This is yet another step to protect data integrity and one that has been done faithfully for the many years that we have been involved in production scanning. In the rare event that the routine detects a photocell that appears to be out of range, we calibrate that machine and perform the test again. If the read is still not up to standard, we call for assistance from our field service engineer.

As a final safeguard, spot checks of scanned files, bubble by bubble and image by image, were routinely made throughout scanning runs. The result of these precautions, from the original layout of the scanning form to the daily vigilance of our operators, was a scan error rate well below .001.

ELECTRONIC DATA FILES

Once the data had been entered and the scanning logs and other paperwork completed, the booklets themselves were put into storage (where they stayed for at least 180 days beyond the close of the fiscal year). When it had been determined that the files were complete and accurate, those files were duplicated electronically and made available for many other processing options. Completed files were loaded onto our local area network (LAN) for transfer to Measured Progress's proprietary I-Score system for scoring. Those files were then used to identify (and print out) papers to be used in the benchmarking and standard-setting processes, and the data made transferable via the Internet, CD-ROM, or optical disk.

ITEMS SCORED BY READERS

Test and answer materials were handled as little as possible to minimize the possibility of loss, mishandling, or breach of security. Once scanned, either by optical mark reader or the I-Score system, papers were stored securely in areas with limited personnel access.

As explained in the following sections on scoring, the I-Score system itself ensures the security of responses and test items: all scoring is “blind”; that is, no student names are associated with viewed responses or raw scores and all scoring personnel are subject to the same nondisclosure requirements and supervision as regular Measured Progress staff.

I-SCORE

After the 2004 test material had been loaded into the LAN, I-Score sent electronically scanned images of student work to individual readers at computer terminals, who evaluated each response and recorded each student’s score via keypad or mouse entry. When the reader had finished with one response, the next response appeared immediately on the computer screen. In that way, the system guaranteed complete anonymity of individual students and ensured the randomization of responses during scoring.

Although I-Score is based on conventional scoring techniques, it also offers numerous benefits, not the least of which is raising the bar on scoring process capability. Some of the benefits follow:

- real-time information on scorer reliability, read-behinds, and overall process monitoring;
- early access to subsets of data for tasks such as standard setting;
- reduced material handling, which not only saves time and labor, but also enhances the security of materials; and
- immediate access to samples of student responses and scores for reporting and analysis through electronic media.

Scoring operations, directed by the manager of scoring services, were carried out by a highly qualified staff. The staff included

- chief readers, who oversaw all training and scoring within particular subject areas;
- quality assurance coordinators (QACs), who led benchmarking and training activities and monitored scoring consistency and rates;
- verifiers, who performed read-behinds of readers and assisted at scoring tables as necessary; and
- readers, who performed the bulk of the scoring.

The table below summarizes the qualifications of the 2003–04 CRT quality assurance coordinators and readers.

2004 Spring Administration					
Scoring Responsibility	Educational Credentials				Total
	Doctorate	Masters	Bachelors	Associates	
QACs	0.00	53.33	46.67	0.00	100%
Readers	4.89	14.66	39.85	40.60	100%

PRELIMINARY ACTIVITIES

Preliminary activities for scoring included (1) participating in the planning and design of documents to be used for scoring, (2) reviewing items and score guides for benchmarking and training and the creation of benchmarking packets, and (3) selecting scoring staff and training them for scoring.

PLANNING AND DESIGNING DOCUMENTS

At the request of the project manager, scoring personnel advised project management and OPI staff on the program design in order to support an efficient and effective scoring process. Scoring staff contributed also to the design of

- response documents and the image-capture process to yield acceptable image clips (also defining file format and layout); and
- scoring benchmarks composed of the guide, subject background information, and anchor papers.

BENCHMARKING

Before the scheduled start of scoring activities, scoring center staff and Montana educators reviewed test items and scoring guides for benchmarking. At that point, chief readers and selected QACs prepared scorer training materials.

Scoring staff from Measured Progress (including test developers) and Montana educators selected one or two anchor examples for each item score point. An additional six to ten responses per item were chosen as part of the training pack. The anchor pack consisted of mid-range exemplars, while the training pack exemplars illustrated the range within each score point. The chief readers, who worked closely with QACs for each content area, facilitated the selection of response exemplars.

SELECTING AND TRAINING SCORING STAFF

QUALITY ASSURANCE COORDINATORS (QACs) AND VERIFIERS

Because the read-behinds performed by the QACs and verifiers moderated the scoring process and thus maintained the integrity of the scores, individuals chosen to fill those positions were selected for their accuracy. In addition, QACs, who train readers to score each item in their content areas, were selected for their ability to instruct and for their level of expertise in their content areas. For this reason, QACs typically are retired teachers who have demonstrated a high level of expertise in their respective disciplines. The ratio of QACs and verifiers to readers was approximately 1:11.

TRAINING QUALITY ASSURANCE COORDINATORS AND VERIFIERS

To ensure that all QACs provided consistent training and feedback, the chief readers spent two days training and qualifying the QACs, and the QACs reviewed all items with the verifiers before scoring. In addition, QACs rotated among tables, supervising readers and reading behind verifiers, who in turn read behind a different table of readers each day.

SELECTING READERS

Applicants were required to demonstrate their ability by participating in a preliminary scoring evaluation. The i-Score system enables Measured Progress to efficiently measure a prospective reader's ability to score student responses accurately. After participating in a training session,

applicants were required to achieve at least 80% exact scoring agreement for a qualifying pack consisting of 20 responses to a predetermined item in their content area. Those 20 responses were randomly selected from a bank of approximately 150, all of which had been selected by QACs and approved by the chief readers and developers.

TRAINING READERS

The QACs first applied the language of the scoring guide for an item to its anchor pack exemplars. Once discussion of the anchor pack had concluded, readers attempted to score the training pack exemplars correctly. The QACs then reviewed the training pack and answered any questions readers had before actual scoring began. With this system, two aspects of scoring efficiency are in conflict. First, in order to minimize training expense, it is desirable to train each reader on as few items as possible. Second, to prevent reader drift and to minimize retraining requirements, it is desirable to score a given item in a brief period of time. But the lower the number of unique items each reader scores, the greater the number of readers required to score that item quickly. To minimize that conflict, we divided each subject area's readers into two or more groups. On the first day of scoring, each group was trained to score a different item. When a group had completed all of an item's responses, those readers were trained on another item (or set).

SCORING ACTIVITIES

Student test booklets at grade level 4 and student response booklets at grade levels 8 and 10 were digitally scanned and scored on a file server for a dedicated, secure LAN. I-Score then distributed digital images of student responses to readers. Training and scoring took place over a period of approximately two weeks.

Items were randomly assigned to readers; thus, each item in a student's response booklet was more than likely scored by a different reader. By using the maximum possible number of readers for each student, the procedure effectively minimized error variance due to reader sampling. All common and matrix-constructed-response items were scored once with a 2% read-behind to ensure consistency among readers and accuracy of individual readers.

MONITORING READERS

After a reader scored a student response, i-Score determined whether that response should also be scored by another reader, scored by a QAC or verifier, or routed for special attention. QACs and verifiers used i-Score to produce daily reader accuracy and speed reports. QACs and verifiers were able to obtain current reader accuracy reports and speed reports on-line at any time.

GENERAL SCORING GUIDES

SHORT-ANSWER ITEMS

Score Point	Description
1	▪ The student's response provides a complete and correct answer.
0	▪ The student's response is totally incorrect or too minimal to evaluate.
B	▪ Blank/no response.

CONSTRUCTED- RESPONSE ITEMS

Score Point	Description
4	<ul style="list-style-type: none">▪ The student completes all important components of the task and communicates ideas clearly.▪ The student demonstrates in-depth understanding of the relevant concepts and/or processes.▪ When instructed to do so, the student chooses more efficient and/or sophisticated processes.▪ When instructed to do so, the student offers insightful interpretations or extensions (e.g., generalizations, applications, and analogies).
3	<ul style="list-style-type: none">▪ The student completes the most important components of the task and communicates clearly.▪ The student demonstrates understanding of major concepts even though he/she overlooks or misunderstands some less important ideas or details.
2	<ul style="list-style-type: none">▪ The student completes most important components of the task and communicates those clearly.▪ The student demonstrates that there are gaps in his/her conceptual understanding.
1	<ul style="list-style-type: none">▪ The student shows minimal understanding.▪ The student addresses <u>only</u> a small portion of the required task(s).
0	▪ The student's response is totally incorrect or irrelevant.
B	▪ Blank/no response.

CHAPTER 8—EQUATING

Equating simply converts raw points from the CRT to a raw score scale. The equated scores then get translated to scaled scores. The process of equating and scaling does not change the rank ordering of students, give more weight to particular questions, or change students' performance-level classifications.

Developing equated scores for the 2004 CRT involved several steps. The first step was to construct the anchor test; that is, to determine the set of equating items. The second step was to calibrate the items in an IRT model. The IRT model used was a combination of the three-parameter logistic (3PL) model for multiple-choice items, the two-parameter logistic (2PL) model for short-answer items, and the graded response model (GRM) for the constructed-response items. The equated scores are translated to the 200 to 300 scale. The following section details the equating process.

GENERAL RULES

- The goal is to have the entire common form be the equating set.
- Equating items cannot be changed from the version used in the previous form in any way.
- Whenever possible, items in the equating set should be selected so that they are within five positions of their location on the previous form.
- Passage sets selected for equating should consist of all, or most, of the items associated with the set.
- The equating set, as a whole, should mirror the characteristics of the common form in terms of content and statistics.

MATH EQUATING DESIGN

Equating items (points) were distributed in approximately the same proportions across strands as the total number of items (points) per strand according to the table on the following page. At least one CR was included in the item set at all grades; at least one short answer item was included at grades 4 and up.

The items were numbered sequentially for each session starting from “1” to determine previous location, not according to how the items were numbered in the operational test booklets. For example, in session two, the matrix items start at location #11 for grades 3-8 and at location #16 for grade 10. Items that are in position 1-5 for the grade 3-8 sessions one and three, and items in position 1-9 on the grade 10 sessions one and three session were not used as equating items.

Strands	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 10
Number & Operations	32 %	32 %	32 %	32 %	30 %	20 %	20 %
Algebra	20 %	20 %	20 %	20 %	20 %	29 %	27 %
Geometry & Measurement	29 %	29 %	29 %	29 %	30 %	32 %	34 %
Data Analysis & Probability	13 %	13 %	13 %	13 %	20 %	20 %	20 %

READING EQUATING DESIGN

Equating items (points) were distributed in approximately the same proportions across strands as the total number of items (points) per strand according to the table on the following page. At least one CR was included in the equating item set at all grades.

The equating item sets for reading were balanced across literary and informational passages. To accomplish this, at least one long passage and two short passages were included in the equating set. If the long passages chosen are informational passages, then two short literary passages were selected and vice versa.

A unique aspect of the augmented forms is that reading matrix items do not start until position #11 for grades 3-8 and position #16 for grade 10. In order to minimize violations of the equating rule that equating items are located within five positions of their position on the previous form, the short passages selected for the equating set became the second passage on the form. Items that are in position 1-5 for the grade 3-8 sessions one and three, and items in position 1-9 on the grade 10 sessions one and three were not used as equating items.

Reading Passage Distribution		
Literary	Stories, poetry, and other forms of literature	50 %
Informational	Content and practical passages	50 %
		100 %
Long	One literary or one informational per session	50 %
Short	At least one literary and informational per session	50 %
		100 %
Strand 1: Comprehension and Analysis		70 %
Strand 2: Reading Process and Language Skills		30 %
		100 %

CHAPTER 9—ITEM ANALYSES

As noted in Brown (1983), “a test is only as good as the items it contains.” A complete evaluation of a test’s quality must include an evaluation of each question. Both the *Standards for Educational and Psychological Testing* (1985) and the *Code of Fair Testing Practices in Education* (1988) include standards for identifying quality questions. Questions should assess only knowledge or skills that are identified as part of the domain being tested and should avoid assessing irrelevant factors. They should also be unambiguous and free of grammatical errors, potentially insensitive content or language, and other confounding characteristics. Further, questions must not unfairly disadvantage test takers from particular racial, ethnic, or gender groups.

Both qualitative and quantitative analyses are conducted to ensure that Montana CRT questions meet these standards. This report focuses on the more quantitative evaluations. The statistical evaluations are presented in three parts: 1) difficulty indices, 2) item-test correlations, and 3) differential item functioning (DIF). The item analyses presented here are based on the statewide administration of the Montana CRT in spring 2004. About 10,600 grade 4 students, 12,100 grade 8 students, and 11,800 grade 10 students participated in the assessment.

DIFFICULTY INDICES (p)

All multiple-choice, constructed-response and short answer items were evaluated in terms of item difficulty according to standard classical test theory practices. Difficulty was defined as the average proportion of points achieved on an item, and was measured by obtaining the average score on an item and dividing by the maximum score for the item. Multiple-choice items were scored dichotomously (correct v. incorrect), so for those questions, the difficulty index is simply the proportion of students who correctly answered the question. Constructed-response items are scored polytomously, where a student can achieve a score of 0, 1, 2, 3, or 4 for the constructed-response items; short answer items (math computation) were scored 0 or 1. By computing the difficulty index as the average proportion of points achieved, the indices for the different item types are placed on a similar scale; the index ranges from 0.0 to 1.0 regardless of the question type. Although this index is traditionally described as a measure of difficulty, it is properly interpreted as an “easiness index” because larger values indicate easier questions.

An index of 0.0 indicates that all students received no credit for the item, and an index of 1.0 indicates that all students received full credit for the item. Ideally, the items on an assessment will have a range of difficulties between 0.25 and 0.9 with most items falling between 0.4 and 0.7. Items that have either a very high or very low difficulty index are considered to be potentially problematic, because they are either so difficult that few students get them right or so easy that nearly all students get them right. In either case, such items should be reviewed for appropriateness for inclusion on the assessment. If an assessment were comprised entirely of very easy or very hard items, all students would receive nearly the same scores and the assessment would not be able to differentiate high-ability students from low-ability students.

ITEM-TEST CORRELATIONS (ITEM DISCRIMINATION)

A desirable feature of an item is that the higher-ability students perform better on the item than lower-ability students. The correlation between student performance on a single item and total test score is a commonly used measure of this characteristic of an item. Within classical test theory, the item-test correlation is referred to as the item's discrimination because it indicates the extent to which successful performance on an item discriminates between high and low scores on the test. The discrimination index used to evaluate Montana CRT multiple-choice and short answer items, which are scored 0 or 1, was the point-biserial correlation between item score and a criterion total score on the test. For constructed-response items, the item discrimination index used was the Pearson product-moment correlation. The theoretical range of these statistics is -1 to $+1$, with a typical range from .3 to .6.

Discrimination indices can be thought of as measures of how closely a question assesses the same knowledge and skills assessed by other questions contributing to the criterion total score. That is, the discrimination index can be thought of as a measure of construct consistency. In light of this interpretation, the selection of an appropriate criterion total score is crucial to the interpretation of the discrimination index. Because each form of the Montana CRT was constructed to be parallel in content, the criterion score selected for each item was the raw score total for each form. The analyses were conducted for each form separately.

SUMMARY OF ITEM ANALYSIS RESULTS

An overall summary of the difficulty and discrimination indices for each grade/content combination is presented in Table 1. The same information broken down by item type (multiple-choice/short-answer, constructed-response, and all items) is shown in Table 2.

The statistics in Table 1 and those calculated for the full set of items in Table 2 are weighted according to the number of points contributed by each item. In an unusual case of a common item with undesirable item statistics, the item is dropped from the operational form.

TABLE 1
ITEM ANALYSIS

Grade	Content Area	Difficulty		Discrimination	
		Mean	StDev	Mean	StDev
4	Reading	0.67	0.14	0.41	0.08
	Mathematics	0.64	0.16	0.34	0.08
8	Reading	0.67	0.15	0.37	0.09
	Mathematics	0.48	0.16	0.32	0.11
10	Reading	0.71	0.16	0.38	0.12
	Mathematics	0.46	0.19	0.31	0.15

TABLE 2
AVERAGE DIFFICULTY AND DISCRIMINATION OF DIFFERENT ITEM TYPES FOR EACH GRADE/CONTENT AREA COMBINATION

Grade	Content Area		Item Type		
			All	MC/SA	Constructed-Response
4	Reading	Difficulty	0.67 (0.14)	0.68 (0.14)	0.47 (0.02)
		Discrimination	0.41 (0.08)	0.41 (0.08)	0.46 (0.01)
		N	54	52	2
	Mathematics	Difficulty	0.64 (0.16)	0.64 (0.16)	0.59 (0.13)
		Discrimination	0.34 (0.08)	0.33 (0.07)	0.45 (0.12)
		N	58	53	5
8	Reading	Difficulty	0.67 (0.15)	0.68 (0.15)	0.49 (0.06)
		Discrimination	0.37 (0.09)	0.36 (0.08)	0.54 (0.03)
		N	54	52	2
	Mathematics	Difficulty	0.48 (0.16)	0.47 (0.16)	0.52 (0.12)
		Discrimination	0.32 (0.11)	0.3 (0.1)	0.5 (0.1)
		N	59	54	5

10	Reading	Difficulty	0.71 (0.16)	0.71 (0.15)	0.51 (0.01)
		Discrimination	0.38 (0.12)	0.37 (0.11)	0.56 (0.05)
		N	59	57	2
	Mathematics	Difficulty	0.46 (0.19)	0.47 (0.19)	0.35 (0.13)
		Discrimination	0.31 (0.15)	0.29 (0.14)	0.51 (0.06)
		N	65	60	5

DIFFERENTIAL ITEM FUNCTIONING (DIF)

Investigations of item or test bias seek to determine whether scores for subgroups of students may be affected by attributes other than those the test is intended to measure. Such investigations usually begin by examining whether subgroups of students performed differently than expected on individual items. Specifically, differences due to irrelevant factors are examined. If such differential item functioning (DIF) is detected, a qualitative assessment of the item is made to determine whether the item is biased against a particular group. It should be noted that the detection of DIF does not imply that the item is biased; instead, it is a statistical tool that helps identify items that may be biased.

The *Code of Fair Testing Practices in Education* (1988) explicitly states that subgroup differences in performance due to irrelevant factors should be examined when sample size permits, and actions should be taken to make certain that differences in performance are due to construct-relevant, rather than irrelevant, factors. The *Standards for Educational and Psychological Testing* includes similar guidelines.

DIF procedures are designed to identify questions for which subgroups of interest perform differentially beyond the impact of differences in overall achievement. For the Montana CRT, the groups that were compared were males and females, white vs. black, and white vs. Hispanic students. For grade 10, the sample size was not large enough to allow statistics to be computed for the white/Hispanic comparison.

The standardization procedure (Dorans and Kulick, 1986) was employed to detect DIF. This procedure calculates the difference in item performance for groups of students matched for achievement on the total test. The standardization index ranges –1 to +1 for multiple-choice items and is adjusted to the

same scale (by dividing by 4) for constructed-response items. Negative numbers indicate that the item was more difficult for female students.

Dorans and Holland (1993) suggested that index values between $-.05$ and $+.05$ should be considered to indicate negligible DIF; values between $-.10$ and $-.05$ and between $+.05$ and $+.10$ be considered to indicate “low” DIF; and that questions with values outside the $[-.10, +.10]$ range be considered to indicate “high” DIF. Each item was categorized according to the guidelines adapted from Dorans and Holland (1993). Tables 3 and 4 provide the number of items in each of the three DIF categories for each grade-content area combination.

TABLE 3
DIF ANALYSIS – ALL GRADES

Grade	Content Area	Male/Female DIF Class									White/Black DIF Class									White/Hispanic DIF Class								
		All			MC/SA			OR			All			MC/SA			OR			All			MC/SA			OR		
		A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
4	Reading	51	2	1	50	1	1	1	1	0	51	5	2	48	3	2	3	2	0	48	9	1	46	6	1	2	3	0
	Math	49	9	0	45	8	0	4	1	0	50	4	0	48	4	0	2	0	0	39	14	1	37	14	1	2	0	0
8	Reading	44	9	1	43	8	1	1	1	0	46	13	0	44	10	0	2	3	0	40	17	2	37	15	2	3	2	0
	Math	47	11	1	43	10	1	4	1	0	49	5	0	47	5	0	2	0	0	43	11	0	41	11	0	2	0	0
10	Reading	52	7	0	52	5	0	0	2	0	50	15	0	46	14	0	4	1	0									
	Math	61	3	1	57	2	1	4	1	0	54	4	1	52	4	1	2	0	0									

A = negligible DIF, B = low DIF, C = high DIF

TABLE 4

**MALE VS. FEMALE DIFFERENTIAL ITEM FUNCTIONING (DIF) CATEGORIZATION BY ITEM TYPE
(MULTIPLE-CHOICE/SHORT-ANSWER AND CONSTRUCTED-RESPONSE)**

Grade	Content Area	Item Type	Negligible DIF				Low DIF				High DIF			
			Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%
4	Reading	MC/SA	28	22	50	96	0	1	1	2	0	1	1	2
		CR	1	0	1	50	1	0	1	50	0	0	0	0
	Math	MC/SA	22	23	45	85	3	5	8	15	0	0	0	0
		CR	4	0	4	80	1	0	1	20	0	0	0	0
8	Reading	MC/SA	27	16	43	83	2	6	8	15	0	1	1	2
		CR	1	0	1	50	1	0	1	50	0	0	0	0
	Math	MC/SA	26	17	43	80	3	7	10	19	0	1	1	2
		CR	4	0	4	80	1	0	1	20	0	0	0	0
10	Reading	MC/SA	28	24	52	91	2	3	5	9	0	0	0	0
		CR	0	0	0	0	2	0	2	100	0	0	0	0
	Math	MC/SA	30	27	57	95	1	1	2	3	0	1	1	2
		CR	4	0	4	80	1	0	1	20	0	0	0	0

ITEM RESPONSE THEORY ANALYSES

In addition to the classical test theory item analyses previously described, the Montana CRT tests were analyzed according to item response theory (IRT) models. IRT uses mathematical models to define a relationship between an unobserved measure of student ability, usually referred to as θ (?), and the probability (p) of getting a dichotomous item correct or of getting a particular score on a polytomous item. In IRT, it is assumed that all items are independent measures of the same construct or ability (i.e., the same ?). There are several commonly used IRT models to specify the relationship between θ and p . For the Montana CRT tests, the 1 parameter logistic (1PL) model was used for multiple-choice and short answer items and the partial credit model was used for the constructed-response items. For more information on IRT and IRT models the reader is referred to Hambleton and van der Linden (1997), and Hambleton and Swaminathan (1985).

The process of determining the specific mathematical relationship between θ and p is referred to as item calibration. Once items are calibrated, they are defined by a set of parameters which specify a non-linear relationship between θ and p . Once the item parameters are known, the θ for each student can be calculated. In IRT, θ is considered to be an estimate of the student's true score or ability and has some characteristics that may make its use preferable to the use of raw scores in rank ordering students

in terms of ability. For more information about item calibration and θ determination the reader is referred to Lord & Novick (1968) or Hambleton & Swaminathan (1985).

For the Montana CRT tests, all items for all forms were simultaneously calibrated. Simultaneously calibrating items across forms has the effect of putting all items on the same scale. Consequently, the θ s for all students will be on the same scale. This important property of IRT analyses was used as the method for equating forms within a year.

PARSCALE v3.5 (Bock & Muraki, 1999) was the software used to do the IRT analyses. The item parameters resulting from the analyses are provided in Section V, Appendix “B”. A one-parameter logistic IRT model was used with the partial credit model being used on the polytomous items. Each item occupied only one block in the calibration run, and the 1.701 normalizing constant was used. A default convergence criterion of 0.001 was used, and all calibrations converged within 30 iterations.

CHAPTER 10—RELIABILITY

Although an individual question's performance is an important focus for evaluation, a complete evaluation of an assessment must also address the way questions function together and complement one another. Tests that function well provide an accurate assessment of the student's level of ability. Unfortunately, no test can do this perfectly. A variety of factors can contribute to a given student's score being either higher or lower than his or her true ability. For example, a student may mis-read a question, or mistakenly bubble in the wrong bubble when he or she knew the answer; similarly a student may get a question correct by guessing, even though he or she did not know the answer. Collectively, these extraneous factors that impact a student's score are referred to as measurement error. Any assessment includes some amount of measurement error; that is, no measurement can be perfectly accurate. This is true of academic assessments—no assessment can measure students perfectly accurately; some students will receive scores that underestimate their true ability, and other students will receive scores that overestimate their true ability. When tests have a high amount of measurement error student scores are very unstable. Students with high ability may get low scores or vice versa. Consequently, one cannot reliably tell a student's true level of ability with such a test. Questions that function well together produce assessments that have less measurement error; that is, the errors made should be small on average and student scores on such a test will consistently represent their ability. Such assessments are described as reliable.

There are a number of ways to estimate an assessment's reliability. One possible approach is to give the same test to the same students at two different points in time. If students receive the same scores on each test, then the extraneous factors affecting performance are small and the test is reliable. (This is referred to as test-retest reliability.) A potential problem with this approach is that students may remember questions from the first administration or may have gained (or lost) knowledge or skills in the interim between the two administrations. A solution to the 'remembering questions' problem is to give a different, but parallel test at the second administration. If student scores on each test correlate highly the test is considered reliable. (This is known as alternate forms reliability, because an alternate form of the test is used in each administration.) This approach, however, does not address the problem that students may have gained (or lost) knowledge or skills in the interim between the two administrations. One way to address these problems is to split the test in half and then correlate students' scores on the two half-tests, this in effect treats each half-test as a complete test. By doing this, the problems associated with an intervening time interval are alleviated. (This is known as a split-

half estimate of reliability.) If the two half-test scores correlate highly, questions on the two half-tests must be measuring very similar knowledge or skills. This is evidence that the questions complement one another and function well as a group. This also suggests that measurement error will be minimal.

The split-half method requires a judgment regarding the selection of which questions contribute to which half-test score. This decision may have an impact on the resulting correlation; different splits will give different estimates of reliability. Cronbach (1951) provided a statistic, α , that avoids this concern about the split-half method. Cronbach's α gives an estimate of the average of all possible splits for a given test. Cronbach's α is often referred to as a measure of internal consistency because it provides a measure of how well all the items in the test measure one single underlying ability.

In addition to Cronbach's α , another approach to estimating the reliability for a test with differing item types (i.e., multiple-choice and constructed response) is to assume that at least a small, but important, degree of unique variance is associated with item type (Feldt and Brennan, 1989). In contrast, Cronbach's coefficient α is built upon the assumption that there are no such local or clustered dependencies. A stratified version of coefficient α corrects for this problem by using the following formula:

$$\alpha_{strat} = 1 - \frac{\sum_{j=1}^k \sigma_{x_j}^2 (1 - \alpha)}{\sigma_x^2}$$

where j indexes the subtests or categories,

$\sigma_{x_j}^2$ represents the variance of the k individual subtests or categories,

α is the unstratified Cronbach's α coefficient, and

σ_x^2 represents the total test variance.

RELIABILITY AND STANDARD ERRORS OF MEASUREMENT

Tables 5 through 7 present Cronbach's α for each test form in each subject area (reading and mathematics), separately for each grade level. The tables also show reliability coefficients separately for multiple-choice/short-answer and constructed-response items, and format stratified reliability coefficients that adjust for the fact that different item formats are included in the test. Table 8 provides descriptive statistics, the overall Cronbach's α coefficient for each grade/content combination, and raw score standard errors of measurement.

TABLE 5
RELIABILITY ANALYSIS – GRADE 4

Content Area	Reliability	Form															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Reading	Coef. α	0.92	0.92	0.92	0.92	0.92	0.91	0.92	0.92	0.91	0.91	0.92	0.92	0.92	0.92	0.92	0.92
	MC/SA α	0.92	0.92	0.92	0.92	0.92	0.91	0.92	0.91	0.91	0.91	0.92	0.92	0.91	0.92	0.92	0.92
	CR α	0.59	0.61	0.62	0.61	0.53	0.58	0.64	0.55	0.63	0.53	0.61	0.60	0.58	0.54	0.52	0.55
	Frmt Strat α	0.92	0.92	0.92	0.92	0.93	0.92	0.93	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
Mathematics	Coef. α	0.89	0.89	0.88	0.89	0.89	0.87	0.90	0.89	0.89	0.89	0.89	0.90	0.89	0.90	0.89	0.89
	MC/SA α	0.87	0.88	0.87	0.88	0.87	0.86	0.89	0.88	0.88	0.88	0.88	0.88	0.87	0.89	0.89	0.88
	CR α	0.61	0.60	0.60	0.62	0.60	0.56	0.62	0.61	0.61	0.58	0.61	0.63	0.62	0.58	0.57	0.61
	Frmt Strat α	0.89	0.90	0.89	0.90	0.89	0.88	0.90	0.89	0.90	0.89	0.90	0.90	0.89	0.90	0.90	0.89

TABLE 6
RELIABILITY ANALYSIS – GRADE 8

Content Area	Reliability	Form															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Reading	Coef. a	0.91	0.90	0.90	0.89	0.90	0.90	0.89	0.90	0.89	0.91	0.90	0.90	0.90	0.89	0.89	0.91
	MC/SA a	0.90	0.90	0.90	0.89	0.89	0.89	0.88	0.89	0.89	0.90	0.90	0.89	0.89	0.89	0.89	0.91
	CR a	0.74	0.75	0.69	0.70	0.74	0.72	0.67	0.69	0.72	0.74	0.70	0.71	0.69	0.74	0.67	0.74
	Frmt Strat a	0.91	0.91	0.91	0.90	0.91	0.91	0.90	0.91	0.90	0.91	0.91	0.91	0.90	0.90	0.90	0.92
Mathematics	Coef. a	0.89	0.88	0.89	0.88	0.88	0.89	0.88	0.88	0.88	0.88	0.88	0.90	0.88	0.88	0.89	0.89
	MC/SA a	0.87	0.86	0.87	0.85	0.85	0.86	0.86	0.86	0.86	0.85	0.86	0.88	0.86	0.86	0.87	0.87
	CR a	0.66	0.66	0.65	0.65	0.65	0.67	0.64	0.65	0.66	0.63	0.64	0.65	0.62	0.63	0.64	0.64
	Frmt Strat a	0.89	0.89	0.89	0.88	0.88	0.89	0.89	0.89	0.89	0.88	0.89	0.90	0.89	0.89	0.89	0.89

TABLE 7
RELIABILITY ANALYSIS – GRADE 10

Content Area	Reliability	Form															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Reading	Coef. a	0.92	0.92	0.91	0.91	0.91	0.91	0.91	0.92	0.91	0.91	0.91	0.92	0.91	0.92	0.91	0.92
	MC/SA a	0.91	0.91	0.91	0.90	0.90	0.90	0.90	0.91	0.90	0.90	0.90	0.91	0.91	0.92	0.90	0.91
	CR a	0.69	0.68	0.63	0.67	0.68	0.65	0.68	0.69	0.66	0.68	0.71	0.69	0.68	0.69	0.66	0.71
	Frmt Strat a	0.92	0.92	0.92	0.91	0.92	0.91	0.91	0.92	0.91	0.91	0.92	0.92	0.92	0.93	0.91	0.93
Mathematics	Coef. a	0.89	0.89	0.89	0.89	0.88	0.90	0.89	0.89	0.89	0.88	0.89	0.89	0.89	0.90	0.88	0.89
	MC/SA a	0.87	0.87	0.86	0.87	0.86	0.88	0.87	0.87	0.87	0.86	0.87	0.87	0.87	0.88	0.85	0.86
	CR a	0.68	0.64	0.67	0.61	0.66	0.66	0.66	0.66	0.67	0.66	0.65	0.65	0.65	0.66	0.64	0.68
	Frmt Strat a	0.90	0.89	0.89	0.89	0.89	0.90	0.89	0.89	0.89	0.88	0.89	0.89	0.89	0.90	0.88	0.89

TABLE 8
RELIABILITIES, STANDARD ERRORS OF MEASUREMENT, AND DESCRIPTIVE STATISTICS

Grade	Content Area	N	Total Points	Mean	SD	Rel	SEM
4	Mathematics	10626	64	40.13	11.06	0.89	3.66
	Reading	10636	60	39.08	10.90	0.92	3.12
8	Mathematics	12093	65	30.73	11.13	0.88	3.78
	Reading	12134	60	39.25	10.20	0.90	3.22
10	Mathematics	11765	71	31.79	11.18	0.89	3.72
	Reading	11796	65	44.68	10.92	0.91	3.22

CHAPTER 11—SCALED SCORES

TRANSLATING RAW SCORES TO SCALED SCORES AND PERFORMANCE LEVELS

Montana CRT scores in each content area are reported on a scale that ranges from 200 to 300. Scaled scores supplement the Montana CRT performance-level results by providing information about the position of a student's results within a performance level. School- and district-level scaled scores are calculated by computing the average of student-level scaled scores. Students' raw scores, or total number of points, on the Montana CRT tests are translated to scaled scores using a data analysis process called **scaling**. Scaling simply converts raw points from one scale to another. In the same way that distance can be expressed in miles or kilometers, or monetary value can be expressed in terms of U.S. dollars or Canadian dollars, student scores on each Montana CRT could be expressed as raw scores (i.e., number right) or scaled scores. It is also important to notice that the raw score to scale score conversion formulae vary from CRT to CRT, analogous to how currency exchange formulae vary from country to country. For example, the scaling conversion formula for Montana's Grade 4 Reading CRT differs from that of the Grade 8 Reading CRT.

It is important to note that converting from raw scores to scaled scores does not change the students' performance-level classifications. Given the relative simplicity of raw scores, it is fair to question why scaled scores are used in Montana CRT reports instead of raw scores. Foremost, scaled scores offer the advantage of simplifying the reporting of results across content areas, grade levels, and subsequent years. Because the standard-setting process typically results in different cut scores across content areas on a raw score basis, it is useful to transform these raw cut scores to a scale that is more easily interpretable and consistent. For the Montana CRT, a score of 225 is the cut score between the **Novice** and **Nearing Proficiency** performance levels. This is true regardless of which content area, grade, or year one may be concerned with. If one were to use raw scores, the raw cut score between **Novice** and **Nearing Proficiency** may be, for example, 35 in mathematics at grade 8, but may be 33 in mathematics at grade 10, or 36 in writing at grade 8. Using scaled scores greatly simplifies the task of understanding how a student performed.

As previously stated, student scores on the Montana CRT are reported in integer values from 200 to 300 with three scores representing cut scores on each assessment. The table on the following page

presents the scaled score range for each performance level in each grade-content area combination. The determination of these cut scores is detailed in the Montana CRT standard setting report.

Grade	Content Area	SCALED SCORE RANGE FOR EACH PERFORMANCE LEVEL			
		Novice	Nearing Proficiency	Proficient	Advanced
4	Reading	200–224	225–249	250–282	283–300
	Mathematics	200–224	225–249	250–285	286–300
8	Reading	200–224	225–249	250–282	283–300
	Mathematics	200–224	225–249	250–293	294–300
10	Reading	200–224	225–249	250–287	288–300
	Mathematics	200–224	225–249	250–287	288–300

The scaled scores are obtained by a simple linear transformation of the Θ s (in this case, raw scores) using the values of 225 and 250 on the scaled score metric and the Θ values obtained through standard setting to define the transformation. For example, the following equation was used to determine the scaled scores for each.

$$SS = m (\Theta) + b$$

where

$$b = 225 - m(\Theta_1)$$

$$m = (225 - 250) / (\Theta_1 - \Theta_2)$$

and SS is the scaled score value, b is the intercept, Θ_1 is the cut score on the Θ metric for the novice/nearing proficiency cut and Θ_2 is the cut score on the Θ metric for the nearing proficiency/proficient cut. In this equation, m represents the slope of the line providing the relationship between Θ and the scaled scores. The scaled score values of 225 and 250 were used because they are the scaled score cut points between novice and nearing proficiency, and nearing proficiency and proficient, respectively. The determination of Θ_1 and Θ_2 is detailed in the Montana CRT standard setting report.

CHAPTER 12—STANDARD SETTING

Standard-Setting Report June 28-29, 2004 Helena, Montana

OVERVIEW OF STANDARD SETTING MEETINGS

The standard-setting meetings to establish cut scores in Reading and Mathematics, Grades 4, 8 and 10, on the Montana CRT were held on Monday June 28th and Tuesday June 29th. There were six panels of 10 to 15 panelists each, and each panel completed the standard-setting activities over the course of the two days.

The standard-setting method implemented for both content areas and all grades is a modified version of the bookmark method. An overview of this method is described below. All panels followed the same procedures.

To help ensure consistency of procedures between panels, each panel was led through the standard setting process by trained facilitators from Measured Progress.

OVERVIEW OF PROCESS

This section of the report provides an overview of the standard-setting process as it was implemented in Montana. The process was divided into the following three stages, each with several constituent tasks.

- ❖ Tasks completed prior to the meeting
 - Creation of performance levels and performance level definitions
 - Preparation of materials for panelists
 - Preparation of presentation materials
 - Preparation of systems and materials for analysis during the meeting
 - Selection of panelists

❖ Tasks completed during the meeting

- Orientation
- Review of assessment materials
- Filling out item map
- Round 1 judgments
- Tabulation of round 1 results
- Round 2: Comparison of panelist judgments and opportunity for revised judgments
- Tabulation of round 2 results
- Round 3: Comparison of panelist results and impact data, and final opportunity to revise judgments
- Modification of Performance Level Definitions
- Evaluation

❖ Tasks completed after the meeting

- Analysis and review of panelists' feedback
- Preparation of recommended cut scores
- Preparation of standard-setting report

TASKS COMPLETED PRIOR TO THE STANDARD SETTING MEETING

CREATION OF PERFORMANCE LEVELS AND PERFORMANCE LEVEL DEFINITIONS

The performance level definitions provided panelists the official description of the knowledge, skills and abilities students are expected to be able to display to be classified into each performance level. These performance level definitions were presented to panelists. The definitions are provided in Appendix "E" of this document.

PREPARATION OF MATERIALS FOR PANELISTS

The following materials were assembled into folders for presentation to the panelists at the standard setting meeting:

- Meeting agenda
- Confidentiality agreement

- Performance level definitions
- Assessment booklet
- Scoring rubrics
- Ordered item booklet
- Item map
- Rating forms
- Evaluation form

PREPARATION OF PRESENTATION MATERIALS

The PowerPoint presentation used in the opening session was prepared prior to the meeting.

PREPARATION OF SYSTEMS AND MATERIALS FOR ANALYSIS DURING THE MEETING

The programming of all analyses to be conducted during the standard-setting meeting was completed and thoroughly tested prior to the standard setting meeting.

SELECTION OF PANELISTS

Panelists were selected prior to the standard-setting meeting. Measured Progress staff selected panelists randomly from among the applicants, and the selected panelists were approved by OPI. Some of the CRT applicants who had special education experience were asked to participate in the CRT-Alternate standard setting instead, and 73 applicants were not selected for either standard setting. The goal was to have 15 panelists for each of the 6 panels, for a total of 90. The actual number of panelists who participated was 75, distributed as follows:

- Grade 4 Math – 15
- Grade 8 Math – 10
- Grade 10 Math – 13
- Grade 4 Reading – 14
- Grade 8 Reading – 13
- Grade 10 Reading – 10

Of the 75 panelists, there were 45 teachers, 18 administrators, and 12 other (parents, librarians, counselors, etc.) Six panelists were Native American, 69 were white, and 51 were female and 24 male. The panelists represented 53 districts in Montana.

TASKS COMPLETED DURING THE STANDARD-SETTING MEETING

The standard-setting meeting on Day 1 began with a general orientation session that was attended by all panelists. The purpose of this session was to provide some background information, provide an introduction to the issues of standard setting, and to explain the activities that would occur during the standard-setting meeting. At the conclusion of the opening session, the floor was opened to questions about the standard-setting process. Most of the questions focused on the uses of the standard-setting results and other policy related issues, although some questions were about the ratings to be made and clarification of the process.

After the large-group session, the panelists assembled into their grade/content area groups. Each group was in a separate room and each room was further divided into two or three tables of 3 to 5 panelists each.

REVIEWING ASSESSMENT MATERIALS

In order to become familiar with the assessment, each panelist took the assessment; the group then scored the assessments together to familiarize themselves with the rubrics.

FILLING OUT ITEM MAP

The purpose of the next step was to ensure that panelists became very familiar with the ordered item booklet and understood the relationships between the ordered items. The ordered item booklet contains one item score category per page, and is ordered from the easiest item category to the most difficult. The ordered item booklet was created by sorting items by their IRT based values (b corresponding to $p^+ = 0.67$ was used). A one-parameter logistic IRT model was used for the dichotomous items and the partial credit IRT model was used for the polytomous items. The room leaders explained to the panelists that each open response item would appear a total of three times in the ordered item booklet, once each for score points 2 through 4.

Each group went through the ordered item booklet, item by item, and discussed the knowledge, skills and competencies students needed to complete each item. Panelists were able to refer to the scoring rubric and the performance level definitions to help them determine this information. Once they were done discussing each item, panelists wrote the knowledge, skills and competencies onto the item map.

ROUND 1 JUDGMENTS

In the first round, subject area panelists worked individually to make their initial judgment of where the bookmarks should be placed. For this task, panelists used the performance level definitions, the item map they completed in the previous step, and the ordered item booklet. Starting with the definition of the Novice performance level, panelists considered the skills and abilities students needed to complete each ordered item and asked themselves the question, “Is a student performing at the Novice level likely to have answered this item correctly?” As they read the items in the booklet, each panelist placed a bookmark (representing the cut score between Novice and Nearing Proficiency) before the first ordered item they felt required skills or knowledge beyond those expected of a student performing at the Novice level. The panelists then repeated this process for the Nearing Proficiency/Proficient and Proficient/Advanced cut scores. Each panelist used the Rating Form provided to record his/her ratings.

TABULATION OF ROUND 1 RESULTS

Each table of panelists received a feedback form that showed where each panelist at the table placed his or her bookmarks, as well as the average for the table. This form was then used to facilitate discussion of the table ratings in round 2.

ROUND 2: COMPARISON OF PANELIST JUDGMENTS AND OPPORTUNITY FOR REVISED JUDGMENTS

During round 2, the panelists at each table examined the results from round 1 and discussed their ratings. The panelists shared their rationale for their bookmark placement in terms of the knowledge and skills students must have to reach that cut score. After all panelists had an opportunity to discuss their bookmark placement and the table completed their discussions, the panelists then had the opportunity to change or revise their round 1 ratings. Each panelist once again used the Rating Form to record his/her ratings.

TABULATION OF ROUND 2 RESULTS

As with round 1, a feedback form was provided to each table after round 2 showing the bookmark placement of each panelist and the average placement for the table. In addition, for round 2 the average placement for the room as a whole was also provided, along with impact data showing the percentage of students that would be placed into each performance level if the room average cut scores were used.

ROUND 3: COMPARISON OF PANELIST RESULTS AND IMPACT DATA, AND FINAL OPPORTUNITY TO REVISE JUDGMENTS

All of the tabulated results from round 2 were distributed to panelists prior to the final round of ratings. As a whole room, panelists discussed the round 2 ratings and the impact data. After the round 3 discussions, each panelist had another opportunity to change or revise his/her ratings, using the Rating Form. The cut scores and percentage of students classified into each performance level, based on the group average cut scores from round 3, were presented to the Technical Advisory Committee and OPI on July 12 & 13, 2004.

MODIFICATION OF PERFORMANCE LEVEL DEFINITIONS

After completing the rating process, the panelists listed suggested modifications to the performance level definitions based on the round 3 results of the standard setting.

EVALUATION

Upon completion of the rating process, panelists anonymously completed an evaluation form.

TASKS COMPLETED AFTER THE STANDARD-SETTING MEETING

Upon conclusion of the standard-setting meeting, several important tasks were completed. These tasks centered on reviewing the standard-setting meeting and addressing anomalies that may have occurred in the process or in the results.

ANALYSIS AND REVIEW OF PANELISTS' FEEDBACK

Upon completion of the evaluation forms, panelists' responses were reviewed and tallied. This review did not reveal any anomalies in the standard-setting process or indicate any reason that a particular panelist's data should not be incorporated in obtaining the final results. Furthermore, it appeared that all panelists understood the rating task and attended to it appropriately (no undue influence was exhibited by a single panelist). An evaluation summary was produced and presented to OPI and the Technical Advisory Committee on July 13, 2004 (see Appendix C: Evaluation Summary).

PREPARE RECOMMENDED CUT SCORES

The recommended cut scores coming out of the standard-setting process are the results from round 3. These cut scores will be reviewed by the Montana Technical Advisory Committee and OPI for approval or adjustment.

PREPARE STANDARD-SETTING REPORT

This report documents the procedures and results of the standard-setting meetings in the establishment of performance standards for the Montana CRT.

TECHNICAL ADVISORY COMMITTEE (TAC)

Montana's Technical Advisory Committee (TAC) met on July 12 & 13, 2004 to review the standard setting selection process, Standard Setting Report, facilitator scripts, panelist's recommended cut scores, and panelist's evaluation summaries. A list of TAC members is included in this report as Appendix D.

TAC members reviewed panelist's judgments plus the standard error of measurement at confidence intervals. NAEP patterns were also reviewed and discussed in depth. The TAC approved the standard-setting process, the Modified Bookmark Method (most commonly used method), and procedures applied by Measured Progress and Montana educators. In addition, the TAC members provided OPI with cut score recommendations. OPI set the final cut scores in all grades and content areas (see Appendix E: CRT Performance Levels and Cut Scores).

Chapter 13—Reporting

The CRT assessments were designed to measure student performance against Montana's Content Standards. Consistent with this purpose, results on the CRT were reported in terms of performance levels that describe student performance in relation to these established state standards. There are four performance levels: *Advanced*, *Proficient*, *Nearing Proficiency*, and *Novice* (CRT Performance Level Descriptors, Scaled Score Ranges, and Raw Scores are described in greater detail in Appendix "E"). Students receive a separate performance-level classification (based on total scaled score) in each content area.

School- and system-level results are reported as the number and percentage of students attaining each performance level at each grade level tested. Disaggregations of students are also reported at the school and system levels. The CRT reports are

- Student Reports;
- Class Roster & Item-Level Reports;
- School Summary Reports; and
- System Summary Reports.

"Decision Rules" were formulated in late spring 2004 by OPI and Measured Progress to identify students, during the reporting process, to be excluded from school and system-level reports. A copy of these "Decision Rules" is included in this report as Appendix "G".

State summary results were provided to OPI on confidential CDs and via a secure Web site. The report formats are included in Appendix "F". These reports were shipped to System Test Coordinators September 20-23, 2004 for distribution to schools within their respective systems/districts. System Test Coordinators and teachers were also provided with copies of the *Guide to Interpretating the 2004 Criterion-Referenced Test and CRT-Alternate Assessment Reports*, to assist them in understanding the connection between the assessment and the classroom. The guide provides information about the assessment and the use of assessment results.

SECTION IV—REFERENCES

- American Educational Research Association (AERA), American Psychological Association, and & National Council on Measurement in Education 1999. *Standards for educational and psychological testing*. Washington, DC: AERA.
- Bock, R.D., and Muraki, E. 1999. *PARSCALE: Parameter scaling of rating data* [Computer program]. Chicago, IL: Scientific Software.
- Brown, F. G. 1983. *Principles of educational and psychological testing* 3rd ed. Fort Worth, TX: Holt, Rinehart, and Winston.
- Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Dorans, N. J., and Holland, P. W. 1993. DIF detection and description. In P. W. Holland and H. Wainer (Eds.), *Differential item functioning* pp. 35–66. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dorans, N. J., and Kulick, E. 1986. Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23, 355–368.
- Hambleton, R. K., and van der Linden, W. J. 1997. *Handbook of modern item response theory*. New York: Springer-Verlag.
- Hambleton, R. K., and Swaminathan, H. 1985. *Item response theory: Principles and applications*. Boston: Kluwer Academic Publishers.
- Joint Committee on Testing Practices 1988. *Code of fair testing practices in education*. Washington, DC: National Council on Measurement in Education.
- Lord, F.M., and Novick, M.R. 1968. *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

SECTION V—APPENDICES

- A. PTS Item and Bias Review Committees & Guidelines for
PTS Reading Passages and Item Bias and Sensitivity
Review**
- B. Item Parameter Files**
- C. Standard Setting: Evaluations**
- D. Technical Advisory Committee**
- E. CRT Performance Level Descriptors, Scaled Scores,
and Raw Scores**
- F. Report Shells**
- G. Reporting Decision Rules**

Appendix A

PTS National Content and Bias Review Committee Members

	Reviewer	Background	Credentials	City	State
1	Barney Berube	ESL	ESL Specialist	Augusta	ME
2	Sheila Murray	Native Am	Wyoming CAC, Grade 4 Teacher	Lander	WY
3	Diep Nguyen	Asian Am	Director of Multicultural Education – SD #54, Illinois	Schaumburg	IL
4	Judith Mounty	Hearing Impaired	Gallaudet University: Teacher	Takoma Park	MD
5	Pamela Mason	African Am	Principal for an elementary school in Milton, MA	Milton	MA
6	Mary Johnson	Native Am	MT Educator	Browning	MT
7	Sarita Kuhn	Native Am		Haver	MT
8	Fern Marx	Gender Equity	Wellesley Center for Women’s Issues	Wellesley	MA
9	Sundra Flansburg	Gender Equity	WEEA Resource Center	Newton	MA
10	Candace Shyer	SPED	NYSED		NY
11	Wendell Bourne	African Am	Director, Cultural Diversity – Cambridge Public Schools	Dorchester	MA
12	Cal Gilbert	Native Am	Principal – Longfellow School	Great Falls	MT
13	Sharif Shakrani	East Indian	Deputy Director – National Assessment Governing Board	Vienna	VA
14	Ethlyn Davis-Fuller	African Am		Brookline	MA
15	Alada Shinault-Small	African Am	College of Charleston – Teacher, Educator	Charleston	SC
16	David Briseno	Hispanic	Director of New Mexico Association of Bilingual Educators	Clovis	NM
17	Roy Howard	Native Am	Western New Mexico University – Bilingual, Teacher, Educator	Gallup	NM
18	Teri Brogdon	SPED	UCHSC	Denver	CO
19	Beverly Chin	Reading	Department of English – University of Montana	Missoula	MT
20	Ann Demers	Reading	Maine Department of Education	Augusta	ME
21	Judy Staten	Reading	Consultant, Item Writer	Rye	NH
22	Karen Walker	Reading	Asst. Superintendent of Instruction – Fountain Valley SD	Fountain Valley	CA
23	Melissa White	Reading	Teacher	Leavittown	PA
24	Judy Coombs	Math	Teacher – George Mitchell School	Waterville	ME
25	Ted Hodgson	Math	Department of Math Sciences – MSU at Bozeman	Bozeman	MT
26	Martin Johnson	Math	University of Maryland	College Park	MD
27	Mary Lindquist	Math	Consultant	Columbus	GA
28	Ken Marks	Math	Director of Math and Science – Tacoma Public Schools	Tacoma	WA

Guidelines
for
Progress Toward Standards
Reading Passage & Item
Bias and Sensitivity Review

Excerpted from a document by
Janice Dowd Scheuneman
Neal Kingston

(Updated by Rachel Slaughter according to suggestions of the
Progress Toward Standards National Bias Committee:
11/17/03)

Bias, Sensitivity and Balance

- **Item Bias**

Item bias stems from item context or content that is irrelevant to the curriculum elements being tested, but affects test scores of an identifiable subgroup of students. For example, several research studies have shown that if you couch a problem intended to test a student's ability to calculate percentages (curricularly relevant) in terms of batting averages (curricularly not relevant) girls will do less well relative to boys than if a non-sports context is used.

- **Sensitivity**

Sensitivity concerns stem from issues that might offend or distract students, but that are not part of the curriculum framework being assessed. Affected students might be identifiable by race, ethnicity or sex, or by more subtle characteristics such as political leanings or religious beliefs.

Sensitivity issues also include situations that might be disturbing to communities based on local events. For example, a reading item about teen suicide might affect the performance of test takers in a school where a student had recently taken his or her life. Sensitive issues are sometimes appropriate as part of instruction, but should be avoided in a test unless required to meet assessment specifications.

- **Curricular Context**

Because both bias and sensitivity concerns must be considered in the context of the curriculum being measured, it is likely that some topics will be appropriate for some subject areas and inappropriate for others. For example, a question on evolution might be appropriate in a science test, or a question about suicide might be appropriate on a health test, but it might be inappropriate to have an item on evolution or health in a reading test.

- **Balance**

Some bias and sensitivity issues arise at the level of test, not item. For example, it is not inappropriate to have a white male or a black female as the character in a question, but it would likely be inappropriate to have all characters in all questions be white males or be black females.

Goals for a Fair and Unbiased Examination

A fair and unbiased examination provides a context that permits all students to demonstrate their achievement and abilities. Students taking an unbiased test should feel that the test is appropriate for them. They should be able to feel that people like themselves are included as part of the assessment activity and are fairly represented in the examination materials. If this goal is to be met, examinations should:

- **Appropriately reflect the diversity of American society**

Items, reading items, essay prompts, and illustrations should present boys and girls, men and women, including those from a variety of racial, ethnic, and language backgrounds, in a non-stereotypic manner.

For example, the content of items should recognize differences of culture among citizens of the United States, e.g., hamburgers and French fries are not typical foods for all cultures in the country and some cultures adhere to a vegetarian diet.

Test contexts should be designed so that they are likely to be familiar to immigrants, newcomers and other groups new to the United States who may also have primary languages other than English. The context of the items should be those that are likely to be the most common possible across these diverse groups. For instance, American baseball and football may not be familiar to students across all cultural groups in the United States. If, however, the passage or item does not require prior knowledge but contains all the information the student needs to answer the question, unfamiliar contexts can be used.

The language of the items should also be reviewed to ensure that it is no more complex than is necessary to assess the knowledge or skill in order to be fair to students for whom English is not the primary language.

- **Use gender-fair language**

If an item does not introduce an individual child or adult, it should be worded to be appropriate for either males or females.

- **Balance the representation of males and females**

Each examination, or discrete part of an examination, should provide a balance of male and female figures.

- **Portray girls, women, and People of Color in active roles**

Women or girls or People of Color should not be represented as passive recipients, observers of actions, or victims in need of rescue by others. If positions of power or status are suggested, the holders of these positions should be balanced in terms of gender with some representation of different racial or ethnic backgrounds.

Portray contemporary women, girls and People of Color as well as historical figures. Portray women, girls and People of Color in ordinary, day-to-day situations, not just as historical figures or extraordinary individuals.

- **Show adults in non-stereotypic professions or work settings**

Adults should be portrayed in ways that reflect the current reality of the workplace. Many positions once considered stereotypically male are now often held by women and People of Color.

- **Distribute positions of power and status among members of different groups**

Power and status should not be portrayed as the exclusive province of a single group. If such positions are used at all, they should belong to both men and women from a variety of backgrounds.

- **Recognize differences among family backgrounds**

No single religious custom or family structure (such as mother, father, and children) should be represented as the norm. As appropriate to the item content, a variety of families should be reflected.

- **Acknowledge the contributions of women and People of Color**

To the extent that items identify the accomplishments of real people, examples should include women and people from various racial, ethnic, and language backgrounds.

- **Portray people with disabilities in a positive manner**

When people with disabilities are portrayed, the material should emphasize their abilities and positive accomplishments rather than emphasize their disabilities.

Fair Tests are Inclusive

Students will generally perform less well if they feel that the testing is for and about others. Presenting testing materials in ways that draw students in will help engage their attention and improve their motivation to perform well on the testing task.

- **All students should feel some connection with the test**

Presentation of a variety of people from different racial/ethnic backgrounds increases the likelihood that students will see people like themselves in the testing situation.

- **Situations presented in the testing materials should be typical for most students or easy for them to imagine**

Students should find it easy to identify with characters in stories or see themselves in the situations presented in test items. To the extent possible, settings should be familiar to all students.

- **Stereotypes serve to distance the student from the material**

Students who are in groups presented in stereotypical situations are given the message that people see their group membership but not their individuality. It suggests that the test is more of the same material developed for somebody else.

- **Items with special appeal for boys or girls or for various racial/ethnic groups should be included**

Research shows that students tend to perform better on items that are of special interest to them. Material should not be allowed to favor the interests of only one group.

- **Language used in items should be inclusive of both men/boys and women/girls**

Avoid the generic he. Find an alternative such as recasting the item in the plural. Other approaches are suggested in the section, *Tips for Avoiding Generic Pronouns*.

Stereotyping

What is a stereotype?

- A standardized picture or mental image
- Oversimplified or exaggerated belief, uncritical judgment
- An unvarying pattern applied to all members of a group
- A lack of recognition of the individuality of the person
- Often accepted as fact
- Stereotypes are not always negative

Stereotypes may be applied to many groups identified by

- Gender
- Race/ethnicity
- National origin
- Religion
- Language or language dialect
- Political affiliation
- Profession
- Area of residence (inner city, rural, suburban)
- Socioeconomic status
- Age
- Sexual orientation
- Physical characteristics (blond, redhead, fat, short, tall)

Dangers of Stereotypes

Stereotypes are not totally irrational and can be a convenient means of coping with diversity. On the other hand, even if they are positive, stereotypes can

- interfere with recognition of an individual's qualities
- reinforce preconceptions of people
- eliminate the need to learn about individuals
- insulate students from real person or group
- affect judgments about people and how they are treated
- reinforce prejudice

- contribute to hostility in relations between groups

Stereotypes may justify believing that

- a group is deserving of a particular fate
- a group is dependent by nature and requires help from other groups (paternalism)
- a group is deficient or lacking in common human attributes such as emotional stability, honesty, industriousness, intelligence, leadership ability, morality, physical appearance, or physical capabilities
- a group is deficient in qualities valued by society such as education, language proficiency, economic condition, political ideology, or professional status

Avoiding Stereotypes

Stereotypes can be well ingrained so that they sound natural and can be easy to miss, particularly those that do not seem negative or may even seem positive in tone.

One test of whether a statement about a racial/ethnic group or about a person from that group is acceptable is to substitute your own group or a member of your group for the one being discussed.

- Statements that seem neutral may be revealed as offensive.
- Statements that appear positive may come across as condescending or paternalistic.
- Statements with negative connotations should, of course, be avoided.

Common Stereotypes

In order to assist in recognizing stereotypes, the following pages list some of the common stereotypes for major population subgroups: Women/girls, African Americans, Asian Americans, and Hispanic Americans. This list is not intended to be exhaustive, but only to illustrate some of the more common stereotypes that might be encountered.

Items that use the stereotypes in the following pages should be amended or deleted if possible.

Stereotypes to be Avoided: Girls/Women

- Overly concerned with physical appearance
- More concerned with home and family than career
- More intuitive, but less logical, than men or boys
- Physically less able than men

- Love to gossip and talk all the time
- All the same, regardless of race and ethnicity
- Spend large amounts of time and money shopping
- Disorganized and scatterbrained
- Emotional and cry easily, at the mercy of their hormones
- Emotions cloud judgment, making them unreliable decision makers
- Not team players
- Lack mechanical abilities and basic mathematics abilities
- Lack leadership qualities such as self-confidence, ambition, or assertiveness
- Less adequately prepared or less competent as professionals

Stereotypes to be Avoided: African American People

- Great athletes, physically powerful
- Musical, great sense of rhythm, terrific entertainers
- Speak “Black” language
- Drive big cars and wear flashy clothes
- Loud, intense, have “attitudes”
- Don’t care about education
- Lazy and shiftless, don’t want to work
- Less adequately prepared or less competent as professionals
- Live in depressed urban areas
- Men often desert their families
- Children have children and become welfare mothers
- Less intelligent than other groups
- Use or sell drugs, carry weapons

Stereotypes to be Avoided: Hispanic American People

- Warm, expressive, and emotional
- Most often work in service or agricultural jobs
- Refuse to learn English
- Don't value education
- Big on machismo, men dominate women
- Lazy and shiftless
- Not punctual and frequently procrastinate
- Don't care if they're on welfare
- Violent and hot tempered

Stereotypes to be Avoided: Asian American People

- Very intelligent, excellent scholars
- Hard working, ambitious, competitive
- Successful in business
- Strong family ties
- Quiet, polite, concerned with proper form
- Inscrutable, concerned with saving face
- Have marriages arranged between families
- Favor sons over daughters
- Prefer to live in ethnic neighborhoods
- Short, skinny, and wear glasses
- Predominantly refugees

Japanese American People

- Law-abiding
- Great imitators

- Sneaky
- Women are servile and obedient

Chinese American People

- Great food
- Run good laundries and restaurants
- Love to gamble
- Use opium or its derivatives
- Cruel

Gender Fair Language

Language that refers to people can be gender-neutral/gender-free or it can be gender-specific.

- **Gender-neutral language is inclusive.**

Gender-neutral language describes people with terms that can be used with either sex. This includes terms such as *student, teacher, writer, player, athlete, and parent*. In recent years, new terms have been introduced to refer to people who were once described by gender-specific terms that implied that job occupants were always men or always women. These new terms include *flight attendant, mail carrier, fire fighter, police officer*.

- Gender-specific terms should not be used to refer to people who may be either males or females.

For many years, it was accepted practice to use masculine pronouns (*he, him, and his*) or the word *man*, as in *mankind*, generically to refer to either males or females. This is no longer the case. If the gender is nonspecific, gender-neutral terms should be used.

Research shows that when *he* or *man* is used to refer to either sex, the majority of people perceive the reference as being to males only.

- **Some terms only appear to be gender-neutral**

Terms such as *doctor, lawyer, politician, minister, and farmer* appear to be gender-neutral, but most people perceive them as men. In order to make these terms refer to women, special efforts may be needed by using a feminine pronoun or a name.

“Dr. Keesha Johnson treated both cats and dogs at her clinic.”

- **Gender-specific language may be appropriate in a gender fair test.**

Gender-specific terms such as *boy/girl, man/woman, mother/father* may promote gender-fair language in situations where use of a gender-neutral term may be interpreted in a gender-specific way.

Examples:

- “*Many children were accompanied by a parent when they had their vaccinations.*”

In this sentence many people might interpret *parent* to be *mother*. In such an instance, it would be better to say that

“*Many children were accompanied by their mother or father when they had their vaccinations.*”

- “*The players on the Spartans softball team traveled to their rival school by bus.*”

With the gender-neutral terms, many people might see the players on the team as boys. Again, the gender-specific terms may actually be more gender-fair.

“*The boys and girls on the Spartans softball team traveled to their rival school by bus.*”

- **Gender-fair language treats males and females equally.**

References to males and females should be symmetric with parallel terms used in the material: *Mr. Smith/Ms. Jones, John Smith/Janet Jones, man/woman, boy/girl, husband/wife.*

Example:

- “*Jorge and his sister each have nine stickers.*”

The girl in this sentence is defined only by her relationship to Jorge. She should at least be named, but probably for a test item her relationship with Jorge is not important and can be omitted.

“*Jorge and Roselia each have nine stickers*”

- **Gender-fair language avoids unwarranted assumptions**

Biased language often treats one type of person, family composition, or way of doing things as the “norm,” implying something deviant or substandard about those who do not conform.

Balance and Equity

Both the individual items and the test, or test section as a whole, need to reflect equivalent treatment of different population subgroups.

- **Gender Balance**

The number of references to males and to females should be nearly the same in subject-matter areas of the test.

- **Balance of Power**

Some figures that are represented in the test items have more power or status than others. In most educational tests, the major power difference will be between child and adult, often a teacher.

Adult figures should also include equal numbers of men and women as well as People of Color. If status differences exist among adult occupations represented, the higher status positions should be distributed among people from different groups.

- **Perspective**

In attempting to convey a variety of environments in which students live, some situations will be more familiar to some students than to others. The items should not be over-balanced toward some settings, such as those that might be more familiar to middle -class families in the suburbs. No one family situation or environment should be presented as the norm.

- **Empowerment**

Woman and People of Color should be portrayed as in control of their lives and destinies and independent of a need for more powerful groups to protect them and fight for their rights.

Areas of Particular Interest for Girls and Boys

Because students tend to do better on materials that interest them, it is advisable to be aware of areas of particular interest.

Some areas that have been identified in research as having differential interest for girls and boys are as follows:

Girls	Boys
Personal relations Aesthetic, philosophic Academic/school concerns Home and family Language, culture	Military, war, weapons Sports Physical Sciences Mechanical, fixing/building things Computers, computer games

Selecting material that may appeal to different interests is appropriate and important. Items likely to be of greater interest to boys or to girls, however, should be balanced in each test form or in each module that will be used to make up a form.

Interest and Prior Knowledge

Material that is interesting to examinees is likely to elicit greater attention to the material and increase motivation to read and understand the question being asked.

Greater interest can also lead to more experience and out-of-school learning about a topic. Care should be taken, therefore, to develop questions that are not made easier for boys or for girls by prior knowledge or experience. If this occurs, the question may actually present an easier task for the group that is more interested in the topic. This therefore would be a biased item.

Sensitive Material

A question that arouses strong emotions in students will most often be inappropriate in an educational achievement test. An emotional response may prevent them from clearly understanding the purpose of the question and the nature of the intended response. In addition, students who become upset during testing will become distracted from the task at hand and may fail to perform as well as they are able.

- **Personal Experience**

If a child has had an experience like that described in an item, will the child be likely to find this upsetting?

Examples: Death in the family, loss of a home

- **Privacy**

Questions should be avoided that may require students to reveal something about themselves or their families that they may not wish to discuss and feel is invasive.

- **Personal Values**

Does the correct response depend on value judgments? This is particularly pertinent when considering how different racial or ethnic groups might respond.

- **Personal Reactions**

Students should not be asked to discuss issues that they may find repugnant or discomforting. For example, students who oppose capital punishment may be distressed if asked to discuss only its merits.

Avoid these topics:

Child Abuse or Neglect	Sexual Orientation
Incest	Occult
Rape	Divorce
Abortion	Parental Conflict
Sex/Sexuality	Suicide

Use these topics with caution:

Death	Family Issues
Guns/Gun Control	Drugs/Alcohol/Tobacco
Homelessness	Murder
Animal Rights	Pregnancy
Racism/Sexism/Ageism	Violence
Religion	Creation/Evolution

Bibliography

American Psychological Association. (1994). Guidelines to Reduce Bias in Language. In Publication Manual of the American Psychological Association, Fourth Edition, pp. 46-60.

Interagency Committee for the Review of the Racial and Ethnic Standards. (July 1997). Recommendations to the Office of Management and Budget Concerning Changes to the Standards for the Classification of Federal Data on Race and Ethnicity. In Federal Register Online [wais.access.gpo.gov], Notices, Volume 62, Number 131, pages 36873-36946.

Maggio, R. (199-). The Bias-Free Word Finder: A Dictionary of Nondiscriminatory Language. Boston: Beacon Press.

National Evaluation Systems, Inc. (1991). Bias Issues in Test Development.

Appendix B: Item Parameter Files

Grade 4 Math

TITLE MONTANA0304 MAT04

>COMMENT ;

MAT04	60	60	7													
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

176197	20001	1.00000	0.00000	-0.30332	0.01273	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176190	20002	1.00000	0.00000	-0.38590	0.01290	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
191242	20003	1.00000	0.00000	-1.54328	0.02072	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176208	20004	1.00000	0.00000	0.46509	0.01306	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176185	20005	1.00000	0.00000	-0.71287	0.01399	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198997	20006	1.00000	0.00000	-0.18924	0.01256	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176254	20007	1.00000	0.00000	-0.56053	0.01339	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176566	20008	1.00000	0.00000	-0.48536	0.01316	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176559	20009	1.00000	0.00000	0.39609	0.01289	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176222	20010	1.00000	0.00000	-1.12785	0.01652	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176184	20011	1.00000	0.00000	-0.66920	0.01380	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176207	20012	1.00000	0.00000	-1.05215	0.01595	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176215	20013	1.00000	0.00000	-0.32321	0.01276	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199001	20014	1.00000	0.00000	-0.45553	0.01307	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176251	20015	1.00000	0.00000	0.15265	0.01251	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176229	20016	1.00000	0.00000	-0.05235	0.01246	0.00000	0.00000
0.00000	0.00000						

0.00000	0.00000						
176562	20017	1.00000	0.00000	-0.62223	0.01362	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176186	20018	1.00000	0.00000	-0.12358	0.01250	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176189	20019	1.00000	0.00000	0.07577	0.01247	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176195	20020	1.00000	0.00000	-0.21318	0.01259	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176572	20021	1.00000	0.00000	-0.45786	0.01308	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199004	20022	1.00000	0.00000	-0.87173	0.01479	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176259	20023	1.00000	0.00000	-0.18817	0.01256	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176241	20024	1.00000	0.00000	-0.63044	0.01365	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166208	20025	1.00000	0.00000	0.13297	0.01250	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199006	20026	1.00000	0.00000	-1.97541	0.02756	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
165023	20027	1.00000	0.00000	0.12183	0.01249	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166362	20028	1.00000	0.00000	-1.34252	0.01845	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166258	20029	1.00000	0.00000	0.33820	0.01277	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
170346	20030	1.00000	0.00000	-0.81836	0.01450	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
165013	20031	1.00000	0.00000	-0.49184	0.01317	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166387	20032	1.00000	0.00000	0.29469	0.01269	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166239	20033	1.00000	0.00000	-0.89690	0.01494	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166355	20034	1.00000	0.00000	-1.56100	0.02095	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199018	20035	1.00000	0.00000	-1.48340	0.01999	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
191244	20036	1.00000	0.00000	-0.07693	0.01247	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						

199019	20037	1.00000	0.00000	-0.24971	0.01264	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176568	20038	1.00000	0.00000	-0.76571	0.01424	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176213	20039	1.00000	0.00000	-0.60809	0.01356	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176249	20040	1.00000	0.00000	-0.27178	0.01267	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178307	20041	1.00000	0.00000	-0.17772	0.01255	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176261	20042	1.00000	0.00000	-0.13554	0.01251	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176211	20043	1.00000	0.00000	-0.85865	0.01472	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176199	20044	1.00000	0.00000	0.45843	0.01304	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176198	20045	1.00000	0.00000	-0.17906	0.01255	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199020	20046	1.00000	0.00000	-0.05948	0.01247	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199021	20047	1.00000	0.00000	0.12952	0.01250	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176196	20048	1.00000	0.00000	-0.73771	0.01410	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199022	20049	1.00000	0.00000	-1.08376	0.01618	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176212	20050	1.00000	0.00000	0.95102	0.01515	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176235	20051	1.00000	0.00000	-0.77331	0.01427	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176210	20052	1.00000	0.00000	0.59085	0.01344	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176217	20053	1.00000	0.00000	-0.79601	0.01439	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176564	20054	1.00000	0.00000	-0.90147	0.01496	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199023	20055	1.00000	0.00000	-0.04469	0.01246	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176204	20056	1.00000	0.00000	-0.84947	0.01467	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						

176203	20057	1.00000	0.00000	-0.44625	0.01305	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176206	20058	1.00000	0.00000	0.16064	0.01252	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176665	50059	1.00000	0.00000	-0.09461	0.00572	0.00000	0.00000
0.00000	0.54548	0.13516	-0.43972	-0.24093			
0.00000	0.02012	0.01756	0.01894	0.01921			
198054	50060	1.00000	0.00000	-0.01113	0.00566	0.00000	0.00000
0.00000	0.41751	0.21091	-0.36172	-0.26669			
0.00000	0.01911	0.01790	0.01889	0.01956			

Grade 8 Math

TITLE MONTANA0304 MAT08

>COMMENT ;

MAT08	60	60	7														
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
191011	20001	1.00000	0.00000	-1.73599	0.02209	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177274	20002	1.00000	0.00000	0.16807	0.01162	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198896	20003	1.00000	0.00000	0.52694	0.01236	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177299	20004	1.00000	0.00000	-0.29452	0.01175	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177324	20005	1.00000	0.00000	0.39499	0.01200	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177334	20006	1.00000	0.00000	-0.83948	0.01356	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177262	20007	1.00000	0.00000	0.25331	0.01173	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198926	20008	1.00000	0.00000	0.17405	0.01163	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177335	20009	1.00000	0.00000	0.27654	0.01177	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
178079	20010	1.00000	0.00000	-0.44384	0.01206	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198927	20011	1.00000	0.00000	0.65464	0.01281	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177345	20012	1.00000	0.00000	0.54573	0.01242	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177333	20013	1.00000	0.00000	0.34959	0.01190	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198928	20014	1.00000	0.00000	0.11297	0.01157	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198929	20015	1.00000	0.00000	0.65995	0.01283	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198930	20016	1.00000	0.00000	0.43323	0.01209	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198931	20017	1.00000	0.00000	0.40063	0.01201	0.00000	0.00000										
0.00000	0.00000																

0.00000	0.00000						
177402	20018	1.00000	0.00000	-0.91288	0.01397	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177322	20019	1.00000	0.00000	0.18049	0.01164	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177266	20020	1.00000	0.00000	0.03039	0.01153	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177403	20021	1.00000	0.00000	0.42108	0.01206	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
191012	20022	1.00000	0.00000	0.36408	0.01193	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177344	20023	1.00000	0.00000	-0.07497	0.01154	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177277	20024	1.00000	0.00000	0.88100	0.01388	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166182	20025	1.00000	0.00000	0.09136	0.01156	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
165827	20026	1.00000	0.00000	0.22441	0.01169	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198932	20027	1.00000	0.00000	-0.18169	0.01161	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166328	20028	1.00000	0.00000	-0.12465	0.01156	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198934	20029	1.00000	0.00000	0.07252	0.01155	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
191013	20030	1.00000	0.00000	-1.00786	0.01455	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
165735	20031	1.00000	0.00000	-0.20533	0.01163	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198935	20032	1.00000	0.00000	-0.09990	0.01155	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
165750	20033	1.00000	0.00000	-0.40923	0.01197	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198937	20034	1.00000	0.00000	0.52720	0.01236	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178085	20035	1.00000	0.00000	-0.74001	0.01308	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177327	20036	1.00000	0.00000	0.43696	0.01210	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177321	20037	1.00000	0.00000	0.42628	0.01208	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						

177342	20038	1.00000	0.00000	0.65519	0.01281	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177300	20039	1.00000	0.00000	0.39132	0.01199	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178441	20040	1.00000	0.00000	0.29518	0.01180	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198938	20041	1.00000	0.00000	0.33902	0.01188	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177297	20042	1.00000	0.00000	0.03695	0.01154	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177320	20043	1.00000	0.00000	-0.40387	0.01196	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177298	20044	1.00000	0.00000	0.00733	0.01153	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177339	20045	1.00000	0.00000	0.19524	0.01165	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178083	20046	1.00000	0.00000	-0.26481	0.01171	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198939	20047	1.00000	0.00000	-0.13943	0.01157	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177323	20048	1.00000	0.00000	0.74127	0.01318	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177182	20049	1.00000	0.00000	-0.09831	0.01155	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177317	20050	1.00000	0.00000	-0.54806	0.01235	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178082	20051	1.00000	0.00000	0.23534	0.01170	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177308	20052	1.00000	0.00000	0.60515	0.01262	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177330	20053	1.00000	0.00000	0.66415	0.01285	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178077	20054	1.00000	0.00000	-0.52818	0.01229	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198940	20055	1.00000	0.00000	0.39768	0.01201	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177332	20056	1.00000	0.00000	-0.62890	0.01263	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177315	20057	1.00000	0.00000	0.05620	0.01154	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						

177310	20058	1.00000	0.00000	0.03017	0.01153	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178091	50059	1.00000	0.00000	0.12728	0.00626	0.00000	0.00000
0.00000	0.64449	0.85841	-0.51026	-0.99263			
0.00000	0.02066	0.01606	0.01506	0.02499			
198069	50060	1.00000	0.00000	0.15123	0.00495	0.00000	0.00000
0.00000	-0.15652	0.58956	-0.68036	0.24733			
0.00000	0.01798	0.01871	0.02180	0.02275			

Grade 10 Math

TITLE MONTANA0304 MAT10

>COMMENT ;

MAT10	65	65	7														
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
175969	20001	1.00000	0.00000	-0.41965	0.01215	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
191062	20002	1.00000	0.00000	0.57394	0.01273	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198766	20003	1.00000	0.00000	0.45055	0.01235	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
175957	20004	1.00000	0.00000	-0.56613	0.01255	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
175833	20005	1.00000	0.00000	0.12156	0.01177	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176002	20006	1.00000	0.00000	0.89494	0.01420	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
175831	20007	1.00000	0.00000	-1.04975	0.01497	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
191063	20008	1.00000	0.00000	0.13853	0.01178	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
166787	20009	1.00000	0.00000	-0.01805	0.01171	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176011	20010	1.00000	0.00000	-0.18031	0.01177	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198767	20011	1.00000	0.00000	-0.06395	0.01171	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
175959	20012	1.00000	0.00000	0.74660	0.01344	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
175807	20013	1.00000	0.00000	0.12745	0.01177	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
175835	20014	1.00000	0.00000	1.16041	0.01601	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198768	20015	1.00000	0.00000	-1.32675	0.01727	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
175806	20016	1.00000	0.00000	0.27175	0.01196	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
175830	20017	1.00000	0.00000	-0.57096	0.01257	0.00000	0.00000										

0.00000	0.00000						
0.00000	0.00000						
175856	20018	1.00000	0.00000	0.06899	0.01173	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175852	20019	1.00000	0.00000	-0.06745	0.01171	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175965	20020	1.00000	0.00000	-0.37510	0.01205	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175964	20021	1.00000	0.00000	1.01767	0.01497	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
191064	20022	1.00000	0.00000	0.77324	0.01356	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175850	20023	1.00000	0.00000	0.15744	0.01180	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175963	20024	1.00000	0.00000	0.34474	0.01209	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198769	20025	1.00000	0.00000	-0.69176	0.01301	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166952	20026	1.00000	0.00000	-0.11905	0.01173	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166150	20027	1.00000	0.00000	0.42733	0.01229	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198770	20028	1.00000	0.00000	-0.58066	0.01260	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166735	20029	1.00000	0.00000	0.04351	0.01172	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166922	20030	1.00000	0.00000	-0.38797	0.01208	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198771	20031	1.00000	0.00000	-0.04135	0.01171	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166139	20032	1.00000	0.00000	-0.51561	0.01240	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166915	20033	1.00000	0.00000	-0.23975	0.01183	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166924	20034	1.00000	0.00000	-0.76578	0.01334	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166956	20035	1.00000	0.00000	0.19379	0.01184	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166967	20036	1.00000	0.00000	-0.71030	0.01309	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166377	20037	1.00000	0.00000	-0.67458	0.01294	0.00000	0.00000

0.00000	0.00000						
0.00000	0.00000						
166940	20038	1.00000	0.00000	-0.40739	0.01212	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
166373	20039	1.00000	0.00000	0.14656	0.01179	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198772	20040	1.00000	0.00000	-0.61994	0.01274	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198773	20041	1.00000	0.00000	0.15649	0.01180	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175828	20042	1.00000	0.00000	0.09054	0.01174	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175951	20043	1.00000	0.00000	-0.23618	0.01183	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175816	20044	1.00000	0.00000	0.62003	0.01290	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175854	20045	1.00000	0.00000	1.66734	0.02140	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198774	20046	1.00000	0.00000	-0.04135	0.01171	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175984	20047	1.00000	0.00000	0.04281	0.01172	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175834	20048	1.00000	0.00000	-0.78006	0.01340	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175829	20049	1.00000	0.00000	0.09500	0.01175	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
191066	20050	1.00000	0.00000	0.12580	0.01177	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176001	20051	1.00000	0.00000	0.30669	0.01202	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175952	20052	1.00000	0.00000	-0.17701	0.01177	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175814	20053	1.00000	0.00000	-0.23760	0.01183	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
191067	20054	1.00000	0.00000	0.55338	0.01266	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175849	20055	1.00000	0.00000	0.57477	0.01274	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175823	20056	1.00000	0.00000	0.95110	0.01454	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175955	20057	1.00000	0.00000	1.21461	0.01646	0.00000	0.00000
0.00000	0.00000						

0.00000	0.00000						
198775	20058	1.00000	0.00000	1.41117	0.01831	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175824	20059	1.00000	0.00000	0.55338	0.01266	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175980	20060	1.00000	0.00000	0.68262	0.01315	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178428	20061	1.00000	0.00000	-0.20817	0.01180	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
175967	20062	1.00000	0.00000	0.54956	0.01265	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178430	20063	1.00000	0.00000	0.46538	0.01239	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176028	50064	1.00000	0.00000	0.60602	0.00667	0.00000	0.00000
0.00000	0.91244	-0.23602	-0.17183	-0.50459			
0.00000	0.01314	0.01784	0.02613	0.03614			
176022	50065	1.00000	0.00000	0.61993	0.00681	0.00000	0.00000
0.00000	0.97031	0.01894	-0.46196	-0.52729			
0.00000	0.01350	0.01583	0.02622	0.03961			

Grade 4 Reading

TITLE MONTANA0304 REA04

>COMMENT ;

REA04	54	54	7														
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
199168	20001	1.00000	0.00000	-1.62409	0.02094	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
199169	20002	1.00000	0.00000	-0.56043	0.01375	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176356	20003	1.00000	0.00000	-0.75265	0.01443	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
199170	20004	1.00000	0.00000	-1.61297	0.02081	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
199171	20005	1.00000	0.00000	-0.92343	0.01523	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176283	20006	1.00000	0.00000	-1.16694	0.01672	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
199172	20007	1.00000	0.00000	-1.06356	0.01603	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176287	20008	1.00000	0.00000	-0.25961	0.01308	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
192348	20009	1.00000	0.00000	-0.66483	0.01409	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176277	20010	1.00000	0.00000	-1.27510	0.01753	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176234	20011	1.00000	0.00000	-0.10973	0.01292	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
184495	20012	1.00000	0.00000	-0.32741	0.01319	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
181028	20013	1.00000	0.00000	-0.95459	0.01539	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176225	20014	1.00000	0.00000	-0.14501	0.01295	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176239	20015	1.00000	0.00000	0.03291	0.01287	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
199173	20016	1.00000	0.00000	-0.83578	0.01479	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
176233	20017	1.00000	0.00000	-0.90696	0.01515	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																

176230	20018	1.00000	0.00000	-0.47858	0.01352	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199174	20019	1.00000	0.00000	-0.78447	0.01456	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
181078	20020	1.00000	0.00000	-0.35713	0.01325	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199175	20021	1.00000	0.00000	0.20619	0.01294	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171026	20022	1.00000	0.00000	-0.23870	0.01306	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171027	20023	1.00000	0.00000	0.18060	0.01292	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171028	20024	1.00000	0.00000	-0.25263	0.01307	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171030	20025	1.00000	0.00000	-0.67431	0.01413	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171029	20026	1.00000	0.00000	-1.00627	0.01569	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
170966	20027	1.00000	0.00000	-1.11032	0.01633	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199176	20028	1.00000	0.00000	-1.21448	0.01707	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
170969	20029	1.00000	0.00000	-1.33863	0.01806	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
170962	20030	1.00000	0.00000	-0.08279	0.01291	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
191254	20031	1.00000	0.00000	-0.10689	0.01292	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
182256	20032	1.00000	0.00000	-0.32386	0.01319	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199177	20033	1.00000	0.00000	0.10139	0.01288	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
184506	20034	1.00000	0.00000	-0.94214	0.01533	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176274	20035	1.00000	0.00000	-1.46234	0.01920	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
181025	20036	1.00000	0.00000	-0.76046	0.01446	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176398	20037	1.00000	0.00000	-0.42805	0.01340	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176374	20038	1.00000	0.00000	-1.15091	0.01661	0.00000	0.00000

0.00000	0.00000						
0.00000	0.00000						
176396	20039	1.00000	0.00000	-1.14622	0.01658	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176375	20040	1.00000	0.00000	-0.81322	0.01469	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199178	20041	1.00000	0.00000	-0.86927	0.01495	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
181086	20042	1.00000	0.00000	-0.50075	0.01358	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
181083	20043	1.00000	0.00000	-0.10519	0.01292	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
181087	20044	1.00000	0.00000	0.03291	0.01287	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199179	20045	1.00000	0.00000	0.10929	0.01288	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
189475	20046	1.00000	0.00000	0.33425	0.01309	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176314	20047	1.00000	0.00000	-1.02651	0.01581	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199180	20048	1.00000	0.00000	-0.51837	0.01363	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199181	20049	1.00000	0.00000	0.14008	0.01290	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176370	20050	1.00000	0.00000	-0.53167	0.01366	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176317	20051	1.00000	0.00000	-0.55081	0.01372	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176318	20052	1.00000	0.00000	-0.34135	0.01322	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
176247	50053	1.00000	0.00000	0.16305	0.00867	0.00000	0.00000
0.00000	1.99783	0.56224	-0.93051	-1.62956			
0.00000	0.03107	0.01441	0.01830	0.03976			
176372	50054	1.00000	0.00000	-0.00149	0.00846	0.00000	0.00000
0.00000	1.95994	0.52057	-1.02272	-1.45779			
0.00000	0.03481	0.01480	0.01778	0.03198			

Grade 8 Reading

TITLE MONTANA0304 REA08

>COMMENT ;

REA08	54	54	7														
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
198105	20001	1.00000	0.00000	-1.90486	0.02392	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198106	20002	1.00000	0.00000	0.32330	0.01204	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177030	20003	1.00000	0.00000	-0.53570	0.01264	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177029	20004	1.00000	0.00000	-0.92367	0.01424	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177026	20005	1.00000	0.00000	-0.14919	0.01190	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177111	20006	1.00000	0.00000	-1.20193	0.01603	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177110	20007	1.00000	0.00000	-1.39103	0.01761	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
181010	20008	1.00000	0.00000	-0.67501	0.01311	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177106	20009	1.00000	0.00000	-0.00674	0.01183	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177103	20010	1.00000	0.00000	-0.47273	0.01247	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177138	20011	1.00000	0.00000	-0.46956	0.01246	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177143	20012	1.00000	0.00000	-1.03044	0.01485	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
184559	20013	1.00000	0.00000	-0.13932	0.01190	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177128	20014	1.00000	0.00000	-0.75093	0.01341	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177129	20015	1.00000	0.00000	-0.60424	0.01286	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177146	20016	1.00000	0.00000	-1.31396	0.01693	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177145	20017	1.00000	0.00000	-1.43148	0.01800	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																

198107	20018	1.00000	0.00000	-0.96716	0.01448	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
180994	20019	1.00000	0.00000	0.00990	0.01182	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177147	20020	1.00000	0.00000	-1.12010	0.01544	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177133	20021	1.00000	0.00000	0.64842	0.01282	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171244	20022	1.00000	0.00000	-1.02969	0.01485	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171245	20023	1.00000	0.00000	-0.65814	0.01305	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171249	20024	1.00000	0.00000	-0.82200	0.01373	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171251	20025	1.00000	0.00000	-0.99274	0.01463	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171253	20026	1.00000	0.00000	-0.29309	0.01209	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171258	20027	1.00000	0.00000	-0.69826	0.01320	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171263	20028	1.00000	0.00000	-0.01506	0.01183	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171265	20029	1.00000	0.00000	-1.55278	0.01925	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
171270	20030	1.00000	0.00000	-0.75216	0.01342	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199389	20031	1.00000	0.00000	-1.12824	0.01549	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198108	20032	1.00000	0.00000	-0.84788	0.01385	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178257	20033	1.00000	0.00000	-0.97288	0.01451	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198109	20034	1.00000	0.00000	0.27429	0.01197	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178261	20035	1.00000	0.00000	-0.53978	0.01265	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
181011	20036	1.00000	0.00000	-0.20974	0.01197	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198110	20037	1.00000	0.00000	-0.31702	0.01213	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177115	20038	1.00000	0.00000	-0.61835	0.01291	0.00000	0.00000

0.00000	0.00000						
0.00000	0.00000						
177123	20039	1.00000	0.00000	-0.90211	0.01413	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177119	20040	1.00000	0.00000	-0.47432	0.01247	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177125	20041	1.00000	0.00000	-0.37735	0.01224	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177032	20042	1.00000	0.00000	-0.25896	0.01203	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177033	20043	1.00000	0.00000	0.45515	0.01229	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198111	20044	1.00000	0.00000	-0.18884	0.01194	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198112	20045	1.00000	0.00000	-0.29110	0.01208	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177035	20046	1.00000	0.00000	-0.76445	0.01347	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177056	20047	1.00000	0.00000	-0.11506	0.01188	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177059	20048	1.00000	0.00000	-0.07487	0.01185	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177044	20049	1.00000	0.00000	-0.36743	0.01222	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198113	20050	1.00000	0.00000	-0.78027	0.01354	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177050	20051	1.00000	0.00000	-0.79248	0.01359	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177061	20052	1.00000	0.00000	0.02773	0.01182	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177148	50053	1.00000	0.00000	-0.17652	0.00731	0.00000	0.00000
0.00000	1.64716	0.50676	-0.74433	-1.40959			
0.00000	0.03455	0.01507	0.01440	0.02344			
177048	50054	1.00000	0.00000	0.22574	0.00669	0.00000	0.00000
0.00000	1.22691	0.42261	-0.53049	-1.11903			
0.00000	0.01955	0.01403	0.01643	0.02767			

Grade 10 Reading

TITLE MONTANA0304 REA10

>COMMENT ;

REA10	59	59	7														
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
198018	20001	1.00000	0.00000	-1.05300	0.01524	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198017	20002	1.00000	0.00000	-1.44655	0.01832	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
178005	20003	1.00000	0.00000	-0.53845	0.01293	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198019	20004	1.00000	0.00000	-0.32747	0.01243	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198020	20005	1.00000	0.00000	-0.32721	0.01243	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198021	20006	1.00000	0.00000	-0.85242	0.01414	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198022	20007	1.00000	0.00000	-0.49525	0.01281	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177947	20008	1.00000	0.00000	-0.46586	0.01273	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
177955	20009	1.00000	0.00000	-0.48885	0.01279	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198023	20010	1.00000	0.00000	-0.32091	0.01242	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
178010	20011	1.00000	0.00000	-0.64014	0.01326	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
178020	20012	1.00000	0.00000	-0.48134	0.01277	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198024	20013	1.00000	0.00000	-0.47193	0.01275	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
178016	20014	1.00000	0.00000	-0.94405	0.01461	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
178030	20015	1.00000	0.00000	-0.98197	0.01482	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198027	20016	1.00000	0.00000	-1.08050	0.01541	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																
198026	20017	1.00000	0.00000	-1.05182	0.01523	0.00000	0.00000										
0.00000	0.00000																
0.00000	0.00000																

198028	20018	1.00000	0.00000	-0.85072	0.01413	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178022	20019	1.00000	0.00000	-0.07195	0.01212	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178017	20020	1.00000	0.00000	-0.62258	0.01320	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178009	20021	1.00000	0.00000	-1.03613	0.01514	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
197602	20022	1.00000	0.00000	-1.79937	0.02242	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
192349	20023	1.00000	0.00000	-0.99511	0.01490	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
194240	20024	1.00000	0.00000	0.48310	0.01258	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
170741	20025	1.00000	0.00000	-0.75337	0.01370	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198030	20026	1.00000	0.00000	-1.16727	0.01600	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
192350	20027	1.00000	0.00000	-1.23273	0.01648	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
192351	20028	1.00000	0.00000	-1.41953	0.01806	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198032	20029	1.00000	0.00000	-1.00953	0.01498	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198033	20030	1.00000	0.00000	-1.00116	0.01493	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
192352	20031	1.00000	0.00000	-0.47996	0.01277	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
192353	20032	1.00000	0.00000	-0.18719	0.01222	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
170830	20033	1.00000	0.00000	-0.84631	0.01411	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198034	20034	1.00000	0.00000	-1.63749	0.02035	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198035	20035	1.00000	0.00000	-1.12496	0.01570	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
192384	20036	1.00000	0.00000	-1.48798	0.01872	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198036	20037	1.00000	0.00000	-1.24341	0.01656	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177925	20038	1.00000	0.00000	-1.33706	0.01732	0.00000	0.00000

0.00000	0.00000						
0.00000	0.00000						
198037	20039	1.00000	0.00000	-1.28481	0.01689	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177921	20040	1.00000	0.00000	-0.62732	0.01322	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198038	20041	1.00000	0.00000	-0.21877	0.01226	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177978	20042	1.00000	0.00000	-1.12454	0.01570	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177981	20043	1.00000	0.00000	-0.77036	0.01377	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177982	20044	1.00000	0.00000	-0.59201	0.01310	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177985	20045	1.00000	0.00000	0.12807	0.01211	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
199239	20046	1.00000	0.00000	-0.62732	0.01322	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198039	20047	1.00000	0.00000	-0.91284	0.01444	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198040	20048	1.00000	0.00000	-0.25880	0.01231	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198041	20049	1.00000	0.00000	-0.89032	0.01433	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198042	20050	1.00000	0.00000	-0.07095	0.01212	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198043	20051	1.00000	0.00000	0.46061	0.01253	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177975	20052	1.00000	0.00000	-0.29116	0.01236	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198045	20053	1.00000	0.00000	-0.11050	0.01215	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198046	20054	1.00000	0.00000	1.70748	0.02075	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
177976	20055	1.00000	0.00000	-0.48440	0.01278	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198047	20056	1.00000	0.00000	-0.90400	0.01440	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
198049	20057	1.00000	0.00000	-0.95243	0.01466	0.00000	0.00000
0.00000	0.00000						
0.00000	0.00000						
178031	50058	1.00000	0.00000	-0.00521	0.00721	0.00000	0.00000
0.00000	1.23888	0.72836	-0.77927	-1.18797			

0.00000	0.02548	0.01566	0.01549	0.02460		
177977	50059	1.00000	0.00000	-0.13235	0.00726	0.00000 0.00000
0.00000	1.52825	0.52718	-0.94251	-1.11293		
0.00000	0.03031	0.01502	0.01597	0.02316		

Appendix C

Standard Setting: Evaluation Summaries

Reading

Grade 4

1. Please mark the subject for which you set standards.

X Reading

☐ Mathematics

2. What was your comfort level with the standard setting process <u>at the beginning</u> of the process?	Extremely Uncomfortable	Somewhat Comfortable	Extremely Comfortable
	1 (1) 2 (6)	3 (6)	4 (1) 5 (0)
3. What was your comfort level with the standard setting process <u>at the end</u> of the process?	Extremely Uncomfortable	Somewhat Comfortable	Extremely Comfortable
	1 (1) 2 (1)	3 (1)	4 (6) 5 (5)
4. To what extent did the training prepare you to complete the task of standard setting?	Not at all	Somewhat Well	Extremely Well
	1 (0) 2 (0)	3 (8)	4 (2) 5 (4)
5. How clear were the performance level definitions?	Not at all Clear	Somewhat Clear	Very Clear
	1 (0) 2 (0)	3 (6)	4 (6) 5 (2)
6. How clear was the bookmarking task?	Not at all Clear	Somewhat Clear	Very Clear
	1 (0) 2 (1)	3 (1)	4 (8) 5 (4)
7. To what extent was the length of this meeting appropriate for the task of setting performance standards?	Too Little Time	About Right	Too Much Time
	1 (0) 2 (0)	3 (11)	4 (2) 5 (1)
8. What is your level of confidence in the bookmarks you placed?	Very Low		Very High
	1 (0) 2 (2)	3 (3)	4 (6) 5 (3)

9. Do you believe the standards set by the panel are correctly placed on the exam score scale?	No		Unsure		Yes
	1 (1)	2 (2)	3 (3)	4 (5)	5 (3)
10. How influential were the following factors in determining the standards you set?	Not at all Influential		Somewhat Influential		Very Influential
A. The performance level descriptors	1 (1)	2 (0)	3 (2)	4 (6)	5 (5)
B. The assessment items	1 (0)	2 (0)	3 (4)	4 (7)	5 (3)
C. Other panelists' comments	1 (0)	2 (1)	3 (6)	4 (5)	5 (2)
D. My professional experience	1 (0)	2 (0)	3 (3)	4 (2)	5 (9)
E. The impact data	1 (0)	2 (3)	3 (3)	4 (6)	5 (2)
F. The political climate	1 (3)	2 (4)	3 (3)	4 (4)	5 (0)
11. Did you find this standard setting session to be professionally rewarding?	No, Not at All		Somewhat		Yes, Extremely
	1 (0)	2 (0)	3 (2)	4 (4)	5 (8)
12. How would you characterize the organization of the standard setting session activities?	Disorganized		Somewhat Organized		Extremely Organized
	1 (0)	2 (0)	3 (2)	4 (7)	5 (5)

13. Please use the space below to provide comments about the standard setting process and/or suggestions as to how the process could be improved.

- It was well-organized and it addressed the information that was needed.
- Very difficult task. Sue was excellent. A valiant attempt to legitimize what is ultimately a very arbitrary and subjective task.
- I would have liked to see the numbers of correct responses for each item.
- There was a lot of down time between rounds 2 and 3 that could have been more beneficially used for round 3. If a second page of chart data does not have relevant info/scores/data, why is

it copied? It seems to be a tremendous waste of resources. For our room, 60 pages of copies were made of second pages that did not contain data on the chart.

- Very important to look at specific skills in each question. This was very enlightening!
- Faster turn around on summarizing data. Save resources (paper) by not printing the second page. Separate classroom teachers from administrators in the small group sessions.
- Too much down time when running numbers. Could have been done in the rooms.
- I am still feeling a disconnect between the item ordering and establishing the cut score. We spent a lot of time determining why items were more difficult than previous items when there were items that were determined as “harder” when they appeared to be easier.
- I’d like to be able to see a print out of how my students did on each items and from that info, maybe adjust my cut marks (I may think they can do it, maybe they didn’t.) Will OPI truly need our input? This needs to be an ongoing process as the test is used. Thanks for giving us exposure to the process.

Reading Grade 8

1. Please mark the subject for which you set standards.

☒ Reading

☐ Mathematics

2. What was your comfort level with the standard setting process <u>at the beginning</u> of the process?	Extremely Uncomfortable		Somewhat Comfortable		Extremely Comfortable
	1 (1)	2 (6)	3 (2)	4 (3)	5 (1)
3. What was your comfort level with the standard setting process <u>at the end</u> of the process?	Extremely Uncomfortable		Somewhat Comfortable		Extremely Comfortable
	1 (0)	2 (0)	3 (2)	4 (6)	5 (5)
4. To what extent did the training prepare you to complete the task of standard setting?	Not at all		Somewhat Well		Extremely Well
	1 (0)	2 (1)	3 (2)	4 (5)	5 (5)
5. How clear were the performance level definitions?	Not at all Clear		Somewhat Clear		Very Clear
	1 (0)	2 (0)	3 (1)	4 (10)	5 (2)
6. How clear was the bookmarking task?	Not at all Clear		Somewhat Clear		Very Clear
	1 (0)	2 (0)	3 (0)	4 (6)	5 (7)
7. To what extent was the length of this meeting appropriate for the task of setting performance standards?	Too Little Time		About Right		Too Much Time
	1 (0)	2 (0)	3 (11)	4 (2)	5 (0)
8. What is your level of confidence in the bookmarks you placed?	Very Low				Very High
	1 (0)	2 (0)	3 (2)	4 (8)	5 (3)
9. Do you believe the standards set by the panel are correctly placed on the exam score scale?	No		Unsure		Yes
	1 (0)	2 (0)	3 (0)	4 (11)	5 (2)

10. How influential were the following factors in determining the standards you set?

	Not at all Influential		Somewhat Influential		Very Influential
A. The performance level descriptors	1 (0)	2 (1)	3 (3)	4 (4)	5 (5)
B. The assessment items	1 (0)	2 (0)	3 (1)	4 (6)	5 (6)
C. Other panelists' comments	1 (0)	2 (0)	3 (4)	4 (4)	5 (5)
D. My professional experience	1 (0)	2 (0)	3 (1)	4 (3)	5 (10)
E. The impact data	1 (1)	2 (2)	3 (6)	4 (2)	5 (2)
F. The political climate	1 (3)	2 (4)	3 (0)	4 (3)	5 (3)

11. Did you find this standard setting session to be professionally rewarding?	No, Not at All		Somewhat		Yes, Extremely
	1 (0)	2 (0)	3 (1)	4 (5)	5 (7)

12. How would you characterize the organization of the standard setting session activities?	Disorganized		Somewhat Organized		Extremely Organized
	1 (0)	2 (0)	3 (0)	4 (4)	5 (9)

13. Please use the space below to provide comments about the standard setting process and/or suggestions as to how the process could be improved.

- I think that it gives a lot of power in setting the bar to very few people.
- It was important to understand background on this process, maybe provide an overview of this process MontCas to date.
- The process is important for all parts of the state of MT to be involved with. I appreciate the knowledge I've attained in this process, a great part of continuing education is learning about where I am going as a certified teacher, this process has helped me accomplish this!
- Thank you for allowing me to be a part of this process. Very interesting and rewarding!
- The facilitator was excellent. The process was clearly defined. The atmosphere here was light, but very task-oriented. Great process.

- I was suspicious of this process, but am now comfortable with our work. Our facilitator was great. She kept us moving and on task! Thanks!
- Susan Izard is a fine facilitator.
- Enjoyable process, knowledgeable staff.
- Facilitator was excellent. Some questions on test need to be looked at #13 and #59
- Process allowed us to discuss in a productive and efficient manner
- I really enjoyed this process and our facilitator, Susan, for grade 8 was awesome! Thank you and I would love to do this again.
- I learned so much! Thank you for all your professionalism and moving us right along.

Reading

Grade 10

1. Please mark the subject for which you set standards.

☒ Reading

☐ Mathematics

2. What was your comfort level with the standard setting process <u>at the beginning</u> of the process?	Extremely Uncomfortable	Somewhat Comfortable	Extremely Comfortable
	1 (3) 2 (5)	3 (1) 4 (1)	5 (0)
3. What was your comfort level with the standard setting process <u>at the end</u> of the process?	Extremely Uncomfortable	Somewhat Comfortable	Extremely Comfortable
	1 (0) 2 (0)	3 (2) 4 (7)	5 (1)
4. To what extent did the training prepare you to complete the task of standard setting?	Not at all	Somewhat Well	Extremely Well
	1 (0) 2 (1)	3 (3) 4 (6)	5 (0)
5. How clear were the performance level definitions?	Not at all Clear	Somewhat Clear	Very Clear
	1 (5) 2 (1)	3 (3) 4 (1)	5 (0)
6. How clear was the bookmarking task?	Not at all Clear	Somewhat Clear	Very Clear
	1 (0) 2 (0)	3 (2) 4 (5)	5 (3)
7. To what extent was the length of this meeting appropriate for the task of setting performance standards?	Too Little Time	About Right	Too Much Time
	1 (1) 2 (0)	3 (8) 4 (1)	5 (0)
8. What is your level of confidence in the bookmarks you placed?	Very Low		Very High
	1 (0) 2 (0)	3 (3) 4 (6)	5 (1)
9. Do you believe the standards set by the panel are correctly placed on the exam score scale?	No	Unsure	Yes
	1 (0) 2 (1)	3 (1) 4 (7)	5 (1)

10. How influential were the following factors in determining the standards you set?

	Not at all Influential		Somewhat Influential		Very Influential
A. The performance level descriptors	1 (1)	2 (1)	3 (4)	4 (2)	5 (2)
B. The assessment items	1 (0)	2 (0)	3 (1)	4 (4)	5 (5)
C. Other panelists' comments	1 (0)	2 (0)	3 (1)	4 (4)	5 (5)
D. My professional experience	1 (0)	2 (0)	3 (0)	4 (5)	5 (5)
E. The impact data	1 (0)	2 (1)	3 (5)	4 (4)	5 (0)
F. The political climate	1 (2)	2 (2)	3 (2)	4 (3)	5 (1)

	No, Not at All		Somewhat		Yes, Extremely
11. Did you find this standard setting session to be professionally rewarding?	1 (0)	2 (0)	3 (0)	4 (2)	5 (8)

	Disorganized		Somewhat Organized		Extremely Organized
12. How would you characterize the organization of the standard setting session activities?	1 (0)	2 (0)	3 (0)	4 (5)	5 (5)

13. Please use the space below to provide comments about the standard setting process and/or suggestions as to how the process could be improved.

- I would like to have had the opportunity to do a “practice” run when we were asked to write out what skills or knowledge each question demanded and how that might cause problems. Not knowing the “busy” words caused trouble under these conditions and writing both columns at the same time for our 6 questions undermined my confidence. The controlled response component caused great difficulty in setting the cut scores! Reading test? Writing test?
- Perhaps more clarity about previously used definitions of levels of student within groups. Added descriptors from the standards or the performance level descriptors will help this. Overall – good process.
- The addition of the constructed response items (which are a writing assessment) was most unfortunate. The items were a distraction and an impediment and created an impasse. The

company, Measured Progress, and its personnel were excellent. A difficult task this was – however, the facilitators and presenters did a fine job.

- I would have liked a better measurement device that was more detail-oriented. I felt as though my experience was more important than a device that should have been driving the standards setting. I also found that the free response questions caused major discussions within our table and room. The fact that they were placed within the questions completely affected the outcome – some placed cuts before and after the written responses. It appeared to us that a reading test should not be subjected to cuts being determined by the writing segment. There were many questions that were poorly written and answers that were misleading. Some of the questions contributed to difficulty of standard setting because of their influence – not because of the reading level. Please re-examine the questions and look carefully.
- The constructed response items were very distracting within the bookmarking process because it was difficult to decide whether we should evaluate the question, the answer, or the text or a combination of the three when evaluating the level of the question. Also, setting cuts for a 10th grade test was extremely difficult to do without having performance level descriptors for the 10th grade. It was encouraging to note that many of us from all over Montana are fairly close to consensus on the skills that can be expected from 10th grade students who are proficient in reading skills.
- I found the process to be valuable. Perhaps we could delineate grade 10 standards more clearly instead of going between grade 8 and grade 12. I hope that some test items are changed – a revision is necessary in some parts.
- I think it became very clear as we went along that at a certain point, people were not going to change their minds. I thought it was smart to have the facilitator jump in and move us on. I think this is an extremely daunting task without 10th grade standards! I wouldn't suggest trying to do this again without grade specific standards to look at!
- When working on bookmark would it be beneficial to divide members of each table for direct discussion as opposed to keeping the tables separated. I would have liked to have ½ our table sit down with ½ the people at the other table.
- Some tasks were not clearly explained. It was a relief to know that we didn't all have to agree on a given cut score, but on the other hand, an average of all of those scores may not really bring us to the “correct” answer either. Overall, while I didn't always agree, I found this a fascinating process!

- Invaluable process – every educator ought to have the opportunity to experience one day session of this process. Given that opportunity, Montana would stop complaining and get focused. I really appreciated working with the professional MP staff. I also appreciated that OPI staff members attended and supported the event. Thank you.

Math GRADE 4

1. Please mark the subject for which you set standards.

☐ Reading

☒ Mathematics

2. What was your comfort level with the standard setting process at the beginning of the process?

Extremely Uncomfortable		Somewhat Comfortable		Extremely Comfortable
1 (1)	2 (4)	3 (6)	4 (4)	5 (0)

3. What was your comfort level with the standard setting process at the end of the process?

Extremely Uncomfortable		Somewhat Comfortable		Extremely Comfortable
1 (0)	2 (0)	3 (2)	4 (8)	5 (5)

4. To what extent did the training prepare you to complete the task of standard setting?

Not at all		Somewhat Well		Extremely Well
1 (0)	2 (1)	3 (4)	4 (8)	5 (2)

5. How clear were the performance level definitions?

Not at all Clear		Somewhat Clear		Very Clear
1 (0)	2 (0)	3 (8)	4 (6)	5 (1)

6. How clear was the bookmarking task?

Not at all Clear		Somewhat Clear		Very Clear
1 (0)	2 (0)	3 (2)	4 (9)	5 (4)

7. To what extent was the length of this meeting appropriate for the task of setting performance standards?

Too Little Time		About Right		Too Much Time
1 (0)	2 (0)	3 (11)	4 (4)	5 (0)

8. What is your level of confidence in the bookmarks you placed?

Very Low				Very High
1 (0)	2 (1)	3 (0)	4 (11)	5 (3)

9. Do you believe the standards set by the panel are correctly placed on the exam score scale?

No		Unsure		Yes
1 (0)	2 (1)	3 (2)	4 (8)	5 (4)

10. How influential were the following factors in determining the standards you set?

	Not at all Influential		Somewhat Influential		Very Influential
A. The performance level descriptors	1 (0)	2 (3)	3 (6)	4 (5)	5 (1)
B. The assessment items	1 (0)	2 (1)	3 (8)	4 (6)	5 (0)
C. Other panelists' comments	1 (0)	2 (1)	3 (7)	4 (3)	5 (4)
D. My professional experience	1 (0)	2 (0)	3 (1)	4 (6)	5 (8)
E. The impact data	1 (0)	2 (2)	3 (4)	4 (6)	5 (2)
F. The political climate	1 (1)	2 (5)	3 (4)	4 (5)	5 (0)

11. Did you find this standard setting session to be professionally rewarding?	No, Not at All		Somewhat		Yes, Extremely
	1 (0)	2 (0)	3 (0)	4 (10)	5 (5)

12. How would you characterize the organization of the standard setting session activities?	Disorganized		Somewhat Organized		Extremely Organized
	1 (0)	2 (0)	3 (0)	4 (11)	5 (4)

13. Please use the space below to provide comments about the standard setting process and/or suggestions as to how the process could be improved.

1. "I was impressed with the professionalism and courtesy exhibited by the Measured Progress staff. This was a good, positive experience. I was frustrated by the amount of time we had to spend waiting for data to be returned so we could go onto the next step. Other than that, thank you so much for all the help and the positive experience."
2. "I'm very concerned that the cut scores will be set high due to our conservation as a group. This test is one area of assessment and should not be anymore important than others. I know the law does not permit it, but the law is wrong! Cut scores should be lower and raised as students and teachers become familiar with the test."
3. "Multiple data input people...could this be somehow electronically input for immediate results. Well organized. Thank you."
4. "This was a very thought provoking experience for me. I am sure it will give me food for thought for many days to come. It was a hard process."
5. "As a classroom teacher I have certainly had apprehensions around the whole CRT testing situation. Being a part of this process has helped dispel some fears and made me feel as though I'm part of this process – it certainly helps!"

6. "It was hard to sit for 2 hours and wait for data to be run. When Chris was "running data" I wish there was something else that we could have been getting done. Thanks – I really enjoyed my time here!"
7. "This would have been a useful process to learn about in college – at least the graduate level. The 2 hour wait for data was a bit frustrating but the overall experience was interesting and will help me in my teaching."
8. "The long wait for data was a waste of time. It would be nice if we had access to the data in the room we were working in. I would also like it if we were notified on how the final cut scores were determined after the other groups look at our recommendations. Thank you."
9. "The only thing I would improve is the lag time that occurred while waiting for data to be crunched! It felt like wasted time. Thanks for the opportunity to do this work!"
10. "I enjoyed this process. I feel it was very beneficial to me and my group. I am interested in seeing how the standards are set."
11. "It was interesting to see/hear and compare answers with my colleagues! It as also an eye-opener to find out where the actual student test scores came out!"
12. "Very interesting process. Gave me a lot to think about. This process really helps to understand the nature of the test and testing."

Math Grade 8

1. Please mark the subject for which you set standards.

☐ Reading

☒ Mathematics

2. What was your comfort level with the standard setting process at the beginning of the process?

Extremely Uncomfortable		Somewhat Comfortable		Extremely Comfortable
1 (1)	2 (3)	3 (4)	4 (0)	5 (2)

3. What was your comfort level with the standard setting process at the end of the process?

Extremely Uncomfortable		Somewhat Comfortable		Extremely Comfortable
1 (1)	2 (0)	3 (0)	4 (4)	5 (5)

4. To what extent did the training prepare you to complete the task of standard setting?

Not at all		Somewhat Well		Extremely Well
1 (0)	2 (0)	3 (1)	4 (3)	5 (6)

5. How clear were the performance level definitions?

Not at all Clear		Somewhat Clear		Very Clear
1 (0)	2 (0)	3 (2)	4 (5)	5 (3)

6. How clear was the bookmarking task?

Not at all Clear		Somewhat Clear		Very Clear
1 (0)	2 (0)	3 (1)	4 (4)	5 (5)

7. To what extent was the length of this meeting appropriate for the task of setting performance standards?

Too Little Time		About Right		Too Much Time
1 (0)	2 (0)	3 (8)	4 (2)	5 (0)

8. What is your level of confidence in the bookmarks you placed?

Very Low				Very High
1 (0)	2 (0)	3 (2)	4 (5)	5 (3)

9. Do you believe the standards set by the panel are correctly placed on the exam score scale?

No		Unsure		Yes
1 (0)	2 (1)	3 (2)	4 (3)	5 (4)

10. How influential were the following factors in determining the standards you set?

	Not at all Influential		Somewhat Influential		Very Influential
A. The performance level descriptors	1 (0)	2 (1)	3 (2)	4 (3)	5 (4)
B. The assessment items	1 (0)	2 (0)	3 (2)	4 (3)	5 (5)
C. Other panelists' comments	1 (0)	2 (0)	3 (2)	4 (5)	5 (3)
D. My professional experience	1 (0)	2 (0)	3 (1)	4 (5)	5 (4)
E. The impact data	1 (0)	2 (1)	3 (6)	4 (2)	5 (1)
F. The political climate	1 (0)	2 (5)	3 (4)	4 (1)	5 (0)

11. Did you find this standard setting session to be professionally rewarding?	No, Not at All		Somewhat		Yes, Extremely
	1 (0)	2 (0)	3 (0)	4 (2)	5 (8)

12. How would you characterize the organization of the standard setting session activities?	Disorganized		Somewhat Organized		Extremely Organized
	1 (0)	2 (0)	3 (1)	4 (4)	5 (5)

13. Please use the space below to provide comments about the standard setting process and/or suggestions as to how the process could be improved.

1. "Overall, I think it all was good. The time frame got bungled up when we had to wait for data to be returned to us. I enjoyed it and learned a lot."
2. "The time lag for round 1 and 2 was too long!"
3. "This was an extremely enlightening experience and very valuable to attend as a teacher. When selecting participants, I feel it is important to select people that have experience (recent) in the classroom and content level."
4. "It was the first time I have experienced book marking. It was an interesting process. I am anxious to see the test results."
5. "I have learned a lot about the CRT and even though I feel the students in my school are at a lower level, it gives me more information to support curriculum changes that are needed in my school."
6. "Great workshop! I feel very good about being part of the process! I loved meeting all the great teachers."
7. "Perhaps some comparison standards from other states would be helpful. The format was well done with plenty of discussion time."

8. “Educational process – I had no expectations ahead of time, so everything we did was a learning experience.”
9. “I feel the proficient standards were too low.”

Math Grade 10

1. Please mark the subject for which you set standards.

☐ Reading

☒ Mathematics

2. What was your comfort level with the standard setting process <u>at the beginning</u> of the process?	Extremely Uncomfortable		Somewhat Comfortable		Extremely Comfortable
	1 (0)	2 (10)	3 (3)	4 (0)	5 (0)
3. What was your comfort level with the standard setting process <u>at the end</u> of the process?	Extremely Uncomfortable		Somewhat Comfortable		Extremely Comfortable
	1 (0)	2 (5)	3 (4)	4 (4)	5 (0)
4. To what extent did the training prepare you to complete the task of standard setting?	Not at all		Somewhat Well		Extremely Well
	1 (0)	2 (1)	3 (6)	4 (4)	5 (1)
5. How clear were the performance level definitions?	Not at all Clear		Somewhat Clear		Very Clear
	1 (0)	2 (2)	3 (6)	4 (5)	5 (0)
6. How clear was the bookmarking task?	Not at all Clear		Somewhat Clear		Very Clear
	1 (0)	2 (1)	3 (2)	4 (3)	5 (7)
7. To what extent was the length of this meeting appropriate for the task of setting performance standards?	Too Little Time		About Right		Too Much Time
	1 (0)	2 (2)	3 (8)	4 (3)	5 (0)
8. What is your level of confidence in the bookmarks you placed?	Very Low				Very High
	1 (1)	2 (4)	3 (2)	4 (6)	5 (0)
9. Do you believe the standards set by the panel are correctly placed on the exam score scale?	No		Unsure		Yes
	1 (2)	2 (1)	3 (2)	4 (6)	5 (2)

10. How influential were the following factors in determining the standards you set?

	Not at all Influential		Somewhat Influential		Very Influential
A. The performance level descriptors	1 (0)	2 (1)	3 (7)	4 (4)	5 (0)
B. The assessment items	1 (0)	2 (0)	3 (3)	4 (6)	5 (4)
C. Other panelists' comments	1 (0)	2 (0)	3 (5)	4 (5)	5 (3)
D. My professional experience	1 (0)	2 (1)	3 (1)	4 (7)	5 (4)
E. The impact data	1 (0)	2 (4)	3 (5)	4 (3)	5 (1)
F. The political climate	1 (0)	2 (2)	3 (8)	4 (0)	5 (3)

	No, Not at All		Somewhat		Yes, Extremely
11. Did you find this standard setting session to be professionally rewarding?	1 (0)	2 (0)	3 (2)	4 (7)	5 (4)

	Disorganized		Somewhat Organized		Extremely Organized
12. How would you characterize the organization of the standard setting session activities?	1 (0)	2 (1)	3 (3)	4 (5)	5 (4)

13. Please use the space below to provide comments about the standard setting process and/or suggestions as to how the process could be improved.

1. "The process was good in general. I am concerned that the level of difficulty was determined by the rate of success on the item. I think that rate was influenced by the curriculum rather than the inherent difficulty of the item."
2. "Measured Progress people did a great job – but this whole process should have been done three years ago. Great job, Scott!"
3. "I would have liked to have had some of the teacher that made up the test (10th grade math) here for input."
4. "I found the process extremely interesting – informative. Thanks for the opportunity."
5. "Time limits for tasks were not clearly stated at start of task. Tasks not explained as completely as they may have been. Not sure of item analysis was worth effort. Would change phase from compare to previous, to "why is this difficult?"
6. "The initial part contains much individual work. While, some is necessary to familiarize ourselves with the test etc., I felt the process became more productive when we discussed the materials as a group."

7. "The process was fine given the position we are in at this point. Concerns: defining easy and hard by number correct; and national policy issues out of our control."
8. "Organize the test questions based on difficulty and not student achievement. Benchmarks would be easier to set and understand why."
9. "I believe that a common distractor could potentially identify some level of proficiency beyond the novice level. We only knew the impact of the correct answer; it was presented in a ranked form only. I would have liked to see the percentages that each choice (a-d) received on each item. It would have provided a clearer picture for me."
10. "It needs to be ongoing as we gather more data. This same process should be revisited every 2 to 3 years."
11. "I was impressed by the process and the dedication of participants to ensure the process was done as expertly as possible. I commend Measured Progress for their execution of the sessions."
12. "I feel that things might have gone smoother if we had a brief history of who and what was done up to this point at the VERY beginning. Many people in our group were filled with misinformation. I was unhappy with the lack of expertise of some of the breakout group members. I was also concerned with how our "random" group was selected.😊"
13. "The process is valued and appropriate. The possible NCLB and AYP have an influence and may have conflicting results on the true purpose of the standard setting exercise. We all have work to do and improving to attempt. Our professional responsibility is enormous in the future 10 years. Thank you for the experience."

Appendix D

Technical Advisory Committee (TAC) Members				
First Name	Last Name	Position	Department	Organization
Art	Bangert, Ph.D.	Assistant Professor	Adult and Higher Education	Montana State University
Sue	Bechard, Ph.D.	Division Director		Measured Progress
Liz	Burton, Ph.D.	Psychometrician	MDA	Measured Progress
Tim	Crockett	Vice President, Client Services		Measured Progress
Carolyn	Haug, Ph.D.	Asst. Division Director		Measured Progress
Sharon	Houle	Program Manager	Client Division	Measured Progress
Michael	Kozlow, Ph.D.	Director		Northwest Regional Ed. Lab
Scott	Marion, Ph.D.	Consultant		Center for Assessment
Mike	Nering, Ph.D.	Psychometrician	MDA	Measured Progress
Madalyn	Quinlan	Chief Executive Officer		OPI
Stanley	Rabinowitz, Ph.D.	Program Director	Assessment & Standards Development Services	WestEd
Nam	Raju, Ph. D.	Distinguished Professor		Institute of Psychology
Steve	Sireci, Ph.D.	Associate Professor		UMASS Amherst
Judy	Snow	State Assessment Director		OPI
Wes	Snyder, Ph.D.	Assistant Vice Pres.	Research & Director of Office of International Programs	University of Montana
Kevin	Sweeney, Ph.D.	Division Director	MDA	Measured Progress
Bud	Williams	Asst. Superintendent		OPI

Appendix E

CRT PERFORMANCE LEVEL DESCRIPTORS

Advanced	This level denotes superior performance.
Proficient	This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
Nearing Proficiency	This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark.
Novice	This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark.

CRT Scaled Score Ranges for Performance Levels

Grade 4

	Reading	Mathematics
Advanced	283-300	286-300
Proficient	250-282	250-285
Nearing Proficiency	225-249	225-249
Novice	200-224	200-224

Grade 8

	Reading	Mathematics
Advanced	283-300	294-300
Proficient	250-282	250-293
Nearing Proficiency	225-249	225-249
Novice	200-224	200-224

Grade 10

	Reading	Mathematics
Advanced	288-300	288-300
Proficient	250-287	250-287
Nearing Proficiency	225-249	225-249
Novice	200-224	200-224

CRT Cut Scores for Performance Levels

TABLE 1

Cut Scores and Impact Data – Grade 4 Reading

Proficiency Level	Minimum Score	% in Level
Advanced	48	26
Proficient	36	39
Nearing Proficiency	27	18
Novice	--	17

TABLE 2

Cut Scores and Impact Data – Grade 8 Reading

Proficiency Level	Minimum Score	% in Level
Advanced	47	27
Proficient	39	31
Nearing Proficiency	33	17
Novice	--	24

TABLE 3

Cut Scores and Impact Data – Grade 10 Reading

Proficiency Level	Minimum Score	% in Level
Advanced	53	31
Proficient	44	32
Nearing Proficiency	38	14
Novice	--	23

TABLE 4

Cut Scores and Impact Data – Grade 4 Math

Proficiency Level	Minimum Score	% in Level
Advanced	53	14
Proficient	43	31
Nearing Proficiency	36	21
Novice	--	34

TABLE 5

Cut Scores and Impact Data – Grade 8 Math

Proficiency Level	Minimum Score	% in Level
Advanced	42	18
Proficient	26	46
Nearing Proficiency	17	25
Novice	--	11

TABLE 6

Cut Scores and Impact Data – Grade 10 Math

Proficiency Level	Minimum Score	% in Level
Advanced	43	18
Proficient	28	42
Nearing Proficiency	18	29
Novice	--	10

Raw-to-Scaled Score Correspondence Grade 4		
Raw Score	Reading Scaled Score	Mathematics Scaled Score
0	200	200
1	200	200
2	200	200
3	200	200
4	200	200
5	200	200
6	200	200
7	200	200
8	200	200
9	200	200
10	200	200
11	200	200
12	200	200
13	200	200
14	200	200
15	200	200
16	200	200
17	200	200
18	200	200
19	203	200
20	206	200
21	208	200
22	211	200
23	214	200
24	217	200
25	219	200
26	222	200
27	225	200
28	228	200
29	231	200
30	233	204
31	236	207
32	239	211
33	242	214
34	244	218
35	247	221
36	250	225
37	253	229
38	256	232
39	258	236
40	261	239
41	264	243
42	267	246
43	269	250
44	272	254
45	275	257
46	278	261

47	281	264
48	283	268
49	286	271
50	289	275
51	292	279
52	294	282
53	297	286
54	300	289
55	300	293
56	300	296
57	300	300
58	300	300
59	300	300
60	300	300
61		300
62		300
63		300
64		300

Raw-to-Scaled Score Correspondence Grade 8		
Raw Score	Reading Scaled Score	Mathematics Scaled Score
0	200	200
1	200	200
2	200	200
3	200	200
4	200	200
5	200	200
6	200	200
7	200	200
8	200	200
9	200	203
10	200	206
11	200	208
12	200	211
13	200	214
14	200	217
15	200	219
16	200	222
17	200	225
18	200	228
19	200	231
20	200	233
21	200	236
22	200	239
23	200	242
24	200	244
25	200	247
26	200	250
27	200	253
28	204	256
29	208	258
30	213	261
31	217	264
32	221	267
33	225	269
34	229	272
35	233	275
36	238	278
37	242	281
38	246	283
39	250	286
40	254	289
41	258	292
42	263	294
43	267	297

44	271	300
45	275	300
46	279	300
47	283	300
48	288	300
49	292	300
50	296	300
51	300	300
52	300	300
53	300	300
54	300	300
55	300	300
56	300	300
57	300	300
58	300	300
59	300	300
60	300	300
61		300
62		300
63		300
64		300
65		300

Raw-to-Scaled Score Correspondence Grade 10		
Raw Score	Reading Scaled Score	Mathematics Scaled Score
0	200	200
1	200	200
2	200	200
3	200	200
4	200	200
5	200	200
6	200	200
7	200	200
8	200	200
9	200	203
10	200	205
11	200	208
12	200	210
13	200	213
14	200	215
15	200	218
16	200	220
17	200	223
18	200	225
19	200	228
20	200	230
21	200	233
22	200	235
23	200	238
24	200	240
25	200	243
26	200	245
27	200	248
28	200	250
29	200	253
30	200	255
31	200	258
32	200	260
33	204	263
34	208	265
35	213	268
36	217	270
37	221	273
38	225	275
39	229	278
40	233	280
41	238	283
42	242	285

43	246	288
44	250	290
45	254	293
46	258	295
47	263	298
48	267	300
49	271	300
50	275	300
51	279	300
52	283	300
53	288	300
54	292	300
55	296	300
56	300	300
57	300	300
58	300	300
59	300	300
60	300	300
61	300	300
62	300	300
63	300	300
64	300	300
65	300	300
66		300
67		300
68		300
69		300
70		300
71		300

Appendix F Report Shells

Student Report

Class Roster & Item-Level Report

School Summary Report

System Summary Report

CRT Performance Level Descriptors

The Performance Level Descriptors below describe students' knowledge, skills, and abilities in a content area. These descriptions provide a picture or profile of student achievement at the four performance levels: advanced, proficient, nearing proficiency, and novice.

Advanced (score to be determined)

This level denotes superior performance.

Proficient (score to be determined)

This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

Nearing Proficiency (score to be determined)

This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark.

Novice (score to be determined)

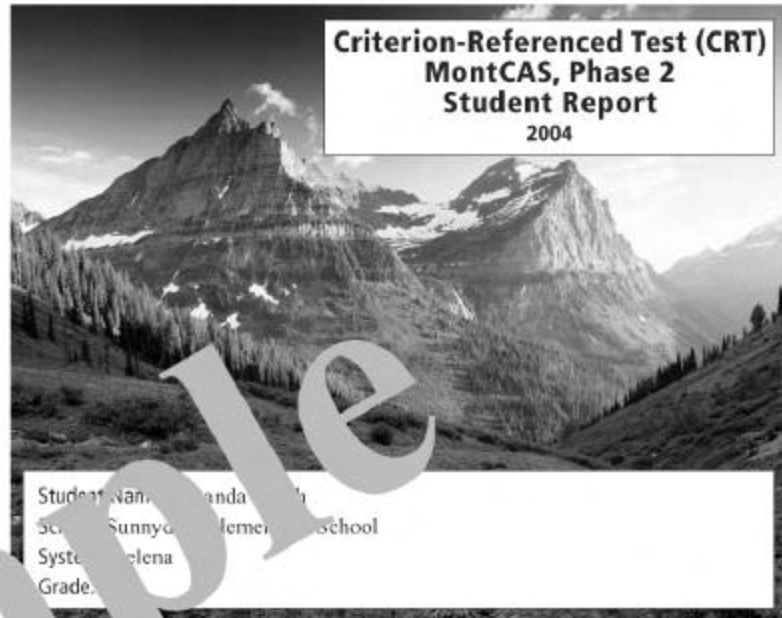
This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark.

OPI Contact
Judy Snow, State Assessment Director
406-444-3656
jsnow@state.mt.us



Linda McCulloch, Superintendent
Montana Office of Public Instruction
PO Box 202501
Helena, Montana 59620-2501
www.opi.state.mt.us

Criterion-Referenced Test (CRT) MontCAS, Phase 2 Student Report 2004



Student Name: [Redacted]
School: Sunnyvale Elementary School
System: Helena
Grade: [Redacted]

Dear Parents/Guardians:

This report contains the results of the first year of the Montana Comprehensive Assessment System Criterion-Referenced Test (CRT) that your child took last spring. The major purpose of the CRT is to provide schools with solid information to evaluate and improve curriculum and instruction to help all students meet Montana's Reading and Mathematics standards. This report provides important information about your child's performance on the assessment, along with state results.

The CRT contains multiple-choice and short-answer questions. The test measures a student's knowledge of subject matter identified in the Montana State Standards for Reading and Mathematics. Your child's results in reading and mathematics are reported in one of four performance levels. These performance levels are defined on the back cover of this report.

It is important to remember that the CRT is just one measure of your child's academic progress. Your local school staff can provide further information about your child's performance in school. The CRT, which is required by the No Child Left Behind Act, is part of an ongoing statewide educational improvement process. Working together, we can ensure that Montana's children continue to receive a high-quality education.

Sincerely,

Linda McCulloch
Montana Superintendent of Public Instruction

How did YOUR CHILD do on the CRT?

Scaled Scores on the CRT

The criterion-referenced test (CRT) is designed to measure student performance against the learning goals described in the Montana Content Standards (<http://www.opi.state.mt.us/standards/index.html>). Consistent with this purpose, results on the CRT are reported according to performance levels that describe student performance in relation to the established state standards. There are four performance levels: **Advanced**, **Proficient**, **Nearing Proficiency**, and **Novice**. Your child's performance levels in reading and mathematics are based on a total scaled score in each content area. Scaled scores in each content area range from 200 to 300. Your child's performance levels, based on the scaled scores, are shown in the bar graphs below.

Scaled Scores

STUDENT RESULTS FOR READING

Performance Level: Novice

Student Scaled Score: 217



STUDENT RESULTS FOR MATHEMATICS

Performance Level: Novice

Student Scaled Score: 217



Scores on Montana Content Standards

In addition to performance levels, CRT results are reported for Montana Content Standards in reading and mathematics. Unlike scaled scores which provide a total performance level score, Montana Content Standard Scores provide more specific information about your child's achievement on the CRT. The chart on the following page shows your child's performance in each area of study within subject areas (Montana Content Standards for reading and math). These results can be used to show your child's relative strengths or weaknesses.

Contact your student's school for more information about the following symbols:

* Student did not complete the assessment.

** Student participated in both CRT and CRT-Alternate. (See additional reports provided.)

§ Student took non-standard accommodation

Scores on Montana Standards

Percentage of Points Earned

Reading Standards

1. Students construct meaning as they comprehend, interpret, and respond to what they read.

Points Possible
15

Student Percentage
7%

State Percentage
10%



2. Students apply a range of skills and strategies to read.

Points Possible
15

Student Percentage
7%

State Percentage
10%

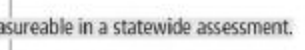


3. Students set goals, monitor, and evaluate their reading progress.

Points Possible
15

Student Percentage
7%

State Percentage
10%



4. Students select, read, and respond to print and non-print materials for a variety of purposes.

Points Possible
15

Student Percentage
7%

State Percentage
10%



5. Students gather, organize, synthesize, and evaluate information from a variety of sources to communicate their findings in ways appropriate for their purposes and audiences.

Points Possible
15

Student Percentage
7%

State Percentage
10%



Math Standards

1. Problem Solving

Points Possible
15

Student Percentage
15%

State Percentage
15%



2. Numbers and Operations

Points Possible
15

Student Percentage
15%

State Percentage
15%



3. Algebra

Points Possible
000

Student Percentage
000%

State Percentage
000%



4. Geometry

Points Possible
000

Student Percentage
000%

State Percentage
000%



5. Measurement

Points Possible
000

Student Percentage
000%

State Percentage
000%



6. Data Analysis, Statistics, and Probability

Points Possible
000

Student Percentage
000%

State Percentage
000%



7. Pattern, Relations, and Functions

Points Possible
000

Student Percentage
000%

State Percentage
000%



○ Student Percentage

■ State Percentage

MontCAS, Phase 2 CRT

School: ELM STREET
System: PORTSMOUTH
Grade: 5
Spring 2004

Reading

School Summary Report

I. Distribution of scores

Perf. Level	Scores	School			System			State		
		N	% of Students	% of Students in Cat.	N	% of Students	% of Students in Cat.	N	% of Students	% of Students in Cat.
Advanced	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX
	XX	XX	XX		XX	XX		XX	XX	
Proficient				XX			XX			XX
Nearing Proficiency										
Novice										

II. Subtest results

Reading	Points Possible	Percentage of Points Earned		
		School	System	State
Total Points	000	000%	000%	000%
Students comprehend texts as they comprehend, interpret, and respond to the texts.	000	000%	000%	000%
Students use a range of strategies to comprehend, interpret, and respond to the texts.				
Students select, read, and respond to print and non-print material for a variety of purposes.				
Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences.				

CRT Performance Level Descriptors

Advanced (score to be determined)

This level denotes superior performance.

Proficient (score to be determined)

This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

Nearing Proficiency (score to be determined)

This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark.

Novice (score to be determined)

This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark.

* Less than ten (10) students were assessed.

MontCAS, Phase 2 CRT

System: PORTSMOUTH
Grade: 5
Spring 2004

Reading

System Summary Report

I. Distribution of scores

Perf. Level	Scores	System			State		
		N	% of Students	% of Students in Cat.	N	% of Students	% of Students in Cat.
Advanced	XX	XX	XX	XX	XX	XX	XX
	XX	XX	XX		XX	XX	
Proficient				XX			XX
Nearing Proficiency							
Novice							

II. Subtest results

Reading	Points Possible	Percentage of Points Earned	
		System	State
Total Points		000%	000%
Students construct meaning as they read, and, interpret and respond to what they read	000	000%	000%
Students use a range of skills and strategies to understand and respond to what they read			
Students select, read, and respond to print and nonprint material for a variety of purposes			
Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences			

CRT Performance Level Descriptors

Advanced (score to be determined)

This level denotes superior performance.

Proficient (score to be determined)

This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

Nearing Proficiency (score to be determined)

This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark.

Novice (score to be determined)

This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark.

* Less than ten (10) students were assessed.

APPENDIX G: MontCAS, Phase 2 CRT
Decision Rules “FINAL”

Participation			Relationship w/ Data File Layouts	Impact on Analyses	Impact on Student report	Impact on School/System/State reports	Impact on Student Roster and I- Analyze	Impact on student level data Excel files for System CD's	Impact on student level data Excel files for State CD
Number of Students ("N")	1	Number of students included in aggregation	NA	N=total number of students with 2 or more responses minus students tested at a private accredited school minus students tested in a non- accredited Title I private school minus foreign exchange students.					
No Class header provided	2	No class indicators provided	Tfname=' ' and Tlname=' '	Class aggregations calculated are actually school level.	No impact	No impact	Report produced	No impact	No impact
Number of Students for Reporting	3	Schools (Systems) has less than 10 students in all content areas	NA	School & Systems with less than 10 students in all content areas is identified. No school/system level data are generated except for item stats on the Student Roster.	No impact	School/system report not Produced	No Impact	No Impact	No Impact
Student Names Not Provided	4	A student answer sheet exists, but no student match is found in the data file.		No Impact on analyses. Student included in DP report to systems. Student counted in N.	No Impact Student name is "Name Not Provided"	Student included	Student Included	Student Included	Student Included

Participation			Relationship w/ Data File Layouts	Impact on Analyses	Impact on Student report	Impact on School/System/State reports	Impact on Student Roster and I- Analyze	Impact on student level data Excel files for System CD's	Impact on student level data Excel files for State CD
Form Number Not Coded	5	DP codes as Form 1, only common items scored		Student counted in N	No Impact	Student Included	No Impact	No Impact	
Tested but Fewer than 2 of the answers marked	6	Student answered fewer than 2 of the common MC questions		Student not counted in N; student excluded from item analysis	Score given with a footnote (†): “Student did not complete the assessment”	Student not Included	Score given with a footnote (†): “Student did not complete the assessment”	Student Included	No Impact
Educational Disability	7	Student has a significant cognitive disability	Disab='1'	If the question “Does the student have a significant cognitive disability?” is coded Yes, student is counted as students with “Significant Cognitive Disability”	No Impact	Counted as student with “Significant Cognitive Disability”	No Impact	No Impact	No Impact
Tested with Standard Accom- modations	8	Student requires an accommodation(s) by content area	Any SA1- SA28 bubbled	If one or more standard accommodations (#1-28) are coded, student is counted as Tested with Standard Accommodation(s)	No Impact	Counted as Tested with Standard Accommodation(s)	No Impact	No Impact	No Impact
Tested with Non- standard	9	Student requires a non-standard accommodation(s)	Any NSA29- NSA32 bubbled	If one or more non-standard accommodations (#29-32) are coded, student is counted as	Student report will indicate raw score with an (§) and a	Student will be given a performance level of “NOVICE” and be	Student record will indicate raw score with an (§) stating	Student will score “NOVICE” and	Student will score “NOVICE”

Participation			Relationship w/ Data File Layouts	Impact on Analyses	Impact on Student report	Impact on School/System/State reports	Impact on Student Roster and I- Analyze	Impact on student level data Excel files for System CD's	Impact on student level data Excel files for State CD
Accom- modations		by content area		Tested with Non-standard Accommodation(s) and will receive a performance level of "NOVICE" and lowest possible scaled score.	footnote stating that the student took a non-standard accommodation.	included in student counts.	that the student took a non-standard accommodation.	be included in student counts.	and be included in student counts.
Program Information	10	Student is identified as participating in an identified program.	SE='1' or Plan504='1' or Migrant='1' or GT='1' or LEP*='1' or Lunch='1' or TM or TR	If one or more Program Information codes are bubbled, student is counted as a program participant	No Impact	Reported on school & system Reporting Category reports.	No Impact	No Impact	No Impact
Special Education- Local Option to Code	11	Student is has an identified disability under IDEA -97.		Student is counted as Special Education; only in CD data files	No Impact	No Impact	No Impact	No Impact	No Impact
Foreign Exchange Student (FXS)	12	Student is identified as a foreign exchange student	Exclusions='1'	Student is not included in any school/system/state aggregations.	Student receives report.	Not included on Reports	Not Included on Reports	Students are not included on System CD	Included on State CD; identified as FXS
Student Not Enrolled (SNE)	13	Student is identified as not enrolled in an accredited public school.	Exclusions='2'	Student is not included in any school/system/state aggregations	Student receives report.	Not Included on Reports	Not Included on Reports	Students are not included on System CD	Included on State CD; identified as SNE

Participation			Relationship w/ Data File Layouts	Impact on Analyses	Impact on Student report	Impact on School/System/State reports	Impact on Student Roster and I- Analyze	Impact on student level data Excel files for System CD's	Impact on student level data Excel files for State CD
Private Accredited School (PRAS)	14	Student is identified as testing at a private accredited school	Exclusions=' 3'	Student is not included in any system/state aggregations	Student receives report.	No school report produced. Not Included on system level reports	Report not produced	Students are not included on System CD	Included on State CD; identified as PRAS
Private Non- accredited Title I School (PRNONST)	15	Student is identified as testing in a non- accredited Title I school	Exclusions=' 4'	Student is not included in any system/state aggregations	Student receives report.	No school report produced. Not Included on system level reports.	Report not produced	Students are not included on System CD	Included on State CD; identified as PRNONST

Did Not Participate in Reading	16	If student is marked as Did Not Participate in Reading and has no responses, student is counted not tested. If student is marked as Did Not Participate in Reading and has responses, the Did Not Participate bubble will be removed in DP.	DNPR	If student did not participate in Reading, student is included in aggregations with minimum scaled score and performance level of Novice	The performance level is given as Novice and minimum scaled score given with footnote (*) “Student did not participate in reading or mathematics.” The points possible for standards are reported. The state average for each standard is reported. The student scores 0s for each standard.	Included with minimum scaled score and performance level of Novice.	Student is included in iAnalyze. Student is included on Roster with footnote (*): “Student did not participate in reading or mathematics.”	Student included on system CD.DNPR coding kept with student record.	Student included on the CD.DNPR coding kept with student record.
Did Not Participate in Math	17	If student is marked as Did Not Participate in Math and has no responses, student is counted as not tested. If student is marked as Did Not Participate in Math and has responses, the Did Not Participate bubble will be removed in DP.	DNPM	If student did not participate in Math, student is included in aggregations with minimum scaled score and performance level of Novice	The performance level is given as Novice and minimum scaled score given with footnote (*) “Student did not participate in reading or mathematics.” The points possible for standards are reported. The state average for each standard is reported. The student scores 0s for each standard.	Included with minimum scaled score and performance level of Novice.	Student is included in iAnalyze. Student is included on Roster with footnote (*): “Student did not participate in reading or mathematics.”	Student included on CD.DNPM coding kept with student record.	Student included on the CD. DNPM coding kept with student record.

Did Not Participate Entire Test	18	If student is marked as Did Not Participate in Entire Test and has no responses, student is counted as not tested. If student is marked as Did Not Participate in Entire Test and has responses, the Did Not Participate bubble will be removed in DP.	DNP	If student did not participate in entire test, student is included in aggregations with minimum scaled score and performance level of Novice for each content area.	The performance level is given as Novice and minimum scaled score given with footnote (**) “Student did not participate in entire test.” The points possible for standards are reported. The state average for each standard is reported. The student scores 0s for each standard.	Included with minimum scaled score and performance level of Novice.	Student is included in iAnalyze. Student is included on Rosters with footnote (**): “Student did not participate in entire test.”	Student included on CD.DNP coding kept with student record.	Student included on CD. DNP coding kept with student record.
CRT-Alternate Only	19	Students in separate data file	ALTR='1' and ALTM='1'	Not included in Analysis	Student receives a CRT-Alternate student report only	Not included in CRT aggregations.	Not included in report	Not included in System CD	Not included in State CD
Participation Information (NSAY & NDAY)	20	Student participated in CRT but has not been a student in school or district for entire academic year	NA	Student is included in participation. If student is marked as NSAY only then student is not included in school aggregations. If student is marked as NDAY then student is not included in either school or district aggregations.	No impact.	If student is marked as NSAY only then student is not included in school data. If student is marked as NDAY then student is not included in school or district data.	If student is NSAY or NDAY student is included on roster with footnote(¥) “Not in school and/or district full academic year.” Student excluded from school (if NSAY or NDAY) and/or district(if NDAY) aggregations.	No Impact	No Impact

Additional Rules:

1. Only common items are used to calculate scores.

*** First year LEP students will be identified in the following way:**

They will have LEP='1' and NSAY='1' and NDAY='1' and have 1 answer marked in the reading test booklet.

Schools: 839

Systems: 281

Scores:

Reading Subtest: Raw score is number of correct responses to common items. Total possible is:

Grade 4: 60 score points

Grade 8: 60 score points

Grade 10: 65 score points

Math Subtest: Raw score is number of correct responses to common items. Total possible is:

Grade 4: 64 score points

Grade 8: 65 score points

Grade 10: 70 score points